ITS
Institut
Teknologi
Sepuluh Nopember

TUGAS AKHIR – IS 184853

PERAMALAN JUMLAH KASUS DEMAM BERDARAH DI KABUPATEN MALANG MENGGUNAKAN METODE RANDOM FOREST

FORECASTING THE AMOUNT OF DENGUE FEVER IN MALANG REGENCY USING RANDOM FOREST

FIRIN HANDAYANI
NRP 05211640000006

Dosen Pembimbing :
Edwin Riksakomara, S.Kom, MT
Raras Tyasnurita, S.Kom., M.BA., Ph.D

DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember
Surabaya 2020

TUGAS AKHIR – IS 184853

PERAMALAN JUMLAH KASUS DEMAM BERDARAH DI KABUPATEN MALANG MENGGUNAKAN METODE RANDOM FOREST

FIRIN HANDAYANI
NRP 05211640000006

Dosen Pembimbing :
Edwin Riksakomara, S.Kom, MT
Raras Tyasnurita, S.Kom., M.BA., Ph.D

DEPARTEMEN SISTEM INFORMASI
Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember
Surabaya 2020

FINAL PROJECT – IS 184853

FORECASTING THE AMOUNT OF DENGUE FEVER IN MALANG REGENCY USING RANDOM FOREST

FIRIN HANDAYANI
NRP 05211640000006

SUPERIVSOR :
Edwin Riksakomara, S.Kom, MT
Raras Tyasnurita, S.Kom., M.BA., Ph.D

DEPARTMENT OF INFORMATION SYSTEMS
Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya 2020

LEMBAR PENGESAHAN

PERAMALAN JUMLAH KASUS DEMAM BERDARAH DI KABUPATEN MALANG MENGGUNAKAN METODE RANDOM FOREST

TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer (S.Kom)
pada
Departemen Sistem Informasi
Fakultas Teknologi Elektro dan Informatika Cerdas (ELECTICS)
Institut Teknologi Sepuluh Nopember

Oleh

Firin Handayani
05211640000006

Surabaya, 14 Agustus 2020

Kepala Departemen Sistem Informasi

Dr. Mudjahidin, ST., MT.
NIP. 197010102003121001

LEMBAR PERSETUJUAN

PERAMALAN JUMLAH KASUS DEMAM BERDARAH DI KABUPATEN MALANG MENGGUNAKAN METODE RANDOM FOREST

TUGAS AKHIR

Disusun Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer
pada
Departemen Sistem Informasi
Fakultas Teknologi Eletro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember

Oleh :

FIRIN HANDAYANI
NRP. 0521164000006

Disetujui Tim Penguji
Tanggal Ujian : 02 Juli 2020
Periode Wisuda : 122

Edwin Riksakomara, S.Kom., MT. **(Pembimbing I)**

Raras Tyasnurita, S.Kom., M.BA., Ph.D. **(Pembimbing II)**

Ahmad Muklason, S.Kom, M.Sc, Ph.D **(Penguji I)**

Faizal Mahananto, S.Kom, M.Eng, Ph.D **(Penguji II)**

PERAMALAN JUMLAH KASUS DEMAM BERDARAH DI KABUPATEN MALANG MENGGUNAKAN METODE RANDOM FOREST

Nama Mahasiswa : Firin Handayani
NRP : 05211640000006
Jurusan : Sistem Informasi FTEIC-ITS
Pembimbing 1 : Edwin Riksakomara, S.Kom, MT
Pembimbing 2 : Raras Tyasnurita, S.Kom., M.BA., Ph.D

ABSTRAK

Demam Berdarah Dengue (DBD) adalah suatu penyakit yang disebabkan oleh nyamuk jenis Aedes Aegypti yang sering terjadi di daerah tropis dan subtropis. Banyak faktor yang dapat memengaruhi perkembangbiakan nyamuk tersebut seperti perubahan suhu udara, kelembapan, curah hujan, mobilitas penduduk, kondisi lingkungan, genangan air, dan perilaku masyarakat yang tidak sehat. Salah satu faktor yang paling besar memengaruhi penularan Virus nyamuk Aedes Aegypti yaitu mobilitas penduduk. Penularan Virus nyamuk Aedes Aegypti akan semakin meningkat sebanding dengan tingkat perubahan kepadatan penduduk.

Penyakit DBD menjadi permasalahan yang sangat serius dalam bidang kesehatan karena dapat menyebabkan kematian. Berdasarkan World Health Organization (WHO), penyakit demam berdarah mengalami peningkatan setiap tahun. Penyakit ini banyak terjadi di negara Asia Tenggara. Selain itu, menurut WHO negara Indonesia termasuk ke dalam delapan negara di Asia yang mengalami peristiwa demam berdarah tertinggi. Jumlah KDB di Indonesia cenderung mengalami peningkatan. Kejadian kasus DBD tertinggi berada di Kabupaten Malang, Jawa Timur. Oleh karena itu, perlu adanya suatu peramalan dimana hasil dari peramalan yang akurat dapat digunakan sebagai informasi untuk pihak Dinas Kesehatan Malang. Selain itu, proses perencanaan

dan pencegahan dapat digunakan dalam menanggulangi jumlah KDBdi masa mendatang.

Metode yang digunakan untuk meramalkan jumlah KDBpada penelitian ini yaitu metode Random Forest Regression. Metode ini dapat digunakan untuk data peramalan time series yang manghasilkan output numerik dengan melakukan berbagai proses pengolahan data.

Model terbaik yang dihasilkan di dataran rendah dan sedang memiliki nilai Root Mean Square Error sebesar 2.2568 dan 2.5964. Sedangkan model terbaik di dataran tinggi memilki nilai Root Mean Square Error paling kecil yaitu 1.3437.

Kata kunci: Demam Berdarah Dengue, Peramalan, Random Forest Regression, Root Mean Square Error.

PERAMALAN JUMLAH KASUS DEMAM BERDARAH DI KABUPATEN MALANG MENGGUNAKAN METODE RANDOM FOREST

Nama Mahasiswa : Firin Handayani
NRP : 05211640000006
Jurusan : Sistem Informasi FTEIC-ITS
Pembimbing 1 : Edwin Riksakomara, S.Kom, MT
Pembimbing 2 : Raras Tyasnurita, S.Kom., M.BA., Ph.D

ABSTRACT

Dengue Fever is a disease caused by Aedes Aegypti mosquitoes that often occur in tropical and subtropical regions. Many factors can influence the breeding of mosquitoes such as changes in temperature, humidity, rainfall, population mobility, environmental conditions, standing water, and unhealthy community behavior. One of the biggest factors influencing the transmission of the Aedes Aegypti mosquito Virus is population mobility. The transmission of the Aedes Aegypti mosquito Virus will increase in proportion to the rate of change in population density.

Dengue Fever becomes a very serious problem in the health field because it can cause death. Based on the World Health Organization (WHO), dengue fever has increased every year. This disease is common in Southeast Asian countries. Also, according to WHO, Indonesia is one of eight countries in Asia with the highest incidence of dengue. The number of dengue cases in Indonesia tends to increase. The highest incidence of Dengue Fever cases is in Malang Regency, East Java. Therefore, there needs to be a forecast where the results of accurate forecasting can be used as information for the Malang Health Office. Also, the planning and prevention process can be used in tackling the number of dengue cases in the future.

The method used to predict the number of dengue cases in this research is the Random Forest Regression method. This method can be used for time series forecasting data which generates numerical output by performing various data processing processes.

The best model produced in the lowlands and medium has a Root Mean Square Error value of 2.2568 and 2.5964. While the best model in the highlands has the smallest Root Mean Square Error value of 1.3437.

Keywords: Dengue Fever, Forecasting, Random Forest Regression, Root Mean Square Error.

SURAT PERNYATAAN BEBAS PLAGIARISM

Saya yang bertandatangan di bawah ini:

Nama	: Firin Handayani
NRP	: 05211640000006
Tempat / Tanggal lahir	: Lumajang / 14 Juni 1997
Fakultas / Departemen	: FTEIC / Sistem Informasi
Nomor Telp / Hp / Email	: 085258193496 / firinhandayani@gmail.com

Dengan ini menyatakan dengan sesungguhnya bahwa penelitian/makalah/tugas akhir saya yang berjudul:

PERAMALAN JUMLAH KASUS DEMAM BERDARAH DI KABUPATEN MALANG MENGGUNAKAN METODE RANDOM FOREST

Bebas Dari Plagiarisme Dan Bukan Hasil Karya Orang Lain.

Apabila dikemudian hari ditemukan seluruh atau sebagian penelitian/makalah/tugas akhir tersebut terdapat indikasi plagiarisme, maka saya bersedia menerima sanksi sesuai peraturan dan ketentuan yang berlaku. Demikian surat pernyataan ini saya buat dengan sesungguhnya dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 08 Agustus 2020

Materai

Firin Handayani

NRP.05211640000006

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji dan syukur penulis tuturkan ke hadirat Allah SWT, Tuhan Semesta Alam yang telah memberikan karunia dan hidayah-Nya kepada penulis sehingga penulis mendapatkan kelancaran dalam menyelesaikan tugas akhir dengan judul:

PERAMALAN JUMLAH KASUS DEMAM BERDARAH DI KABUPATEN MALANG MENGGUNAKAN METODE RANDOM FOREST

Terima kasih penulis sampaikan kepada pihak-pihak yang telah mendukung, memberikan motivasi, semangat, saran dan bantuan baik berupa material maupun moril demi tercapainya tujuan pembuatan tugas akhir ini. Tugas akhir ini tidak akan pernah terwujud tanpa bantuan dan dukungan dari berbagai pihak yang sudah meluangkan waktu, tenaga dan pikirannya. Secara khusus penulis akan menyampaikan ucapan terima kasih yang sebanyak-banyaknya kepada :

1. Ibu Seniyem selaku ibu kandung dari penulis yang tiada henti memberikan dukungan, motivasi, dan semangat dari awal perkuliahan sampai menyelesaikan tugas akhir.
2. Bapak Edwin Riksakomara, S.Kom., MT. selaku dosen pembimbing 1 dan Ibu Raras Tyasnurita, S.Kom., M.BA., Ph.D. selaku dosen pembimbing 2 yang senantiasa meluangkan waktu untuk berdiskusi, memberikan ilmu dan saran, serta memotivasi dalam kelancaran pengerjaan tugas akhir.
3. Ibu Wiwik Anggraeni, S.Si., M.Kom. selaku dosen yang memberikan izin untuk menggunakan data dari penelitian beliau dan memberikan saran serta motivasi untuk kelancaran pengerjaan tugas akhir.
4. Bapak Ahmad Muklason, S.Kom., M.Sc., Ph.D. dan Faisal Mahananto, S.Kom, M.Eng, Ph.D. selaku dosen penguji yang

telah memberikan saran dan kritik yang sangat membangun untuk perbaikan tugas akhir.

5. Seluruh dosen Jurusan Sistem Informasi ITS yang telah memberikan ilmu yang bermanfaat kepada penulis.
6. Humaira dan Berta selaku sahabat dan partner dalam mengerjakan tugas akhir serta memberikan dukungan untuk segera menyelesaikan tugas akhir.
7. Fika, Meli, Humayun, dan Lulu' yang selalu memebrikan semangat dan canda tawa serta motivasi dalam proses pengerjaan tugas akhir.
8. M. Reza Pahlawan dan M. Wildan Maulidani yang telah membantu dalam memahami materi dan kode selama pengerjaan tugas akhir.
9. Teman-teman kabinet HMSI Rumah Karya khususnya Departemen *Social Development* dan teman-teman seperjuangan dalam lab RDIB yang telah banyak memberikan saran dan semangat selama masa perkuliahan.
10. Rekan-rekan ARTEMIS yang telah memberikan banyak kenangan selama masa perkuliahan.
11. Berbagai pihak yang tidak bisa disebutkan satu persatu yang telah turut serta menyukseskan penulis dalam menyelesaikan tugas akhir.

Penyusunan laporan ini masih jauh dari kata sempurna sehingga penulis menerima adanya kritik maupun saran yang membangun untuk perbaikan di masa yang akan datang. Semoga buku tugas akhir ini dapat memberikan manfaat bagi pembaca.

Surabaya, 08 Agustus 2020

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR KODE

BAB I
PENDAHULUAN

Pada bab pendahuluan menjelaskan tentang latar belakang masalah, perumusan masalah, batasan masalah, tujuan tugas akhir, manfaat tugas akhir, dan relevansi tugas akhir dengan bidang keilmuan sistem informasi. Berdasarkan uraian pada bab ini diharapkan memberi suatu gambaran umum tentang permasalahan dan penyelesaian pada tugas akhir.

1.1. Latar Belakang

Demam berdarah *dengue* merupakan suatu penyakit yang ditularkan oleh nyamuk jenis *Aedes aegypti* yang dapat menyebabkan berbagai macam penyakit. Sebanyak 50-100 juta kasus penyakit demam berdarah terjadi di daerah tropis dan sub-tropis setiap tahun dengan jumlah kematian sekitar 20.000 jiwa. Penyakit demam berdarah *dengue* termasuk permasalahan yang sangat penting dalam bidang kesehatan yang terjadi di negara Asia Tenggara [1].

Penyakit DBD disebabkan oleh nyamuk *Aedes aegypti* dimana terdapat beberapa faktor yang memengaruhi tingkat perkembangbiakan dari nyamuk tersebut. Faktor-faktor tersebut yaitu curah hujan yang tinggi, kondisi lingkungan lembab, adanya genangan air, sanitasi lingkungan yang buruk, perilaku masyarakat yang tidak sehat, dan mobilitas penduduk. Penularan Virus nyamuk *Aedes aegypti* paling besar dipengaruhi oleh adanya mobilitas penduduk dengan mengikuti pertumbuhan kepadatan penduduk [2].

Jumlah penderita demam berdarah *dengue* di Asia berada pada urutan pertama menurut data dari seluruh dunia. Berdasarkan hasil *World Health Organization* (WHO), penyakit demam berdarah mengalami peningkatan setiap tahun. Pada tahun 1990-1997 terdapat 479.848 kejadian demam berdarah *dengue* sedangkan pada tahun 2000-2007 terdapat 925.896. Selain itu, menurut WHO negara Indonesia termasuk ke dalam delapan negara di Asia yang mengalami kasus DBD tertinggi. Negara-

negara tersebut yaitu Indonesia, Myanmar, Bangladesh, India, Maldives, Sri Lanka, Thailand, dan Timor Leste [3].

Berdasarkan data Kementerian Kesehatan, kasus DBD pada pertengahan Januari 2019 sudah tercatat sebanyak 4.798 kasus dimana kasus DBD didominasi terjadi di pulau Jawa. Jumlah KDBsebanyak 690 kasus terjadi di Jawa Tengah, 541 kasus di Jawa Barat, dan 1.081 kasus di Jawa Timur. Kejadian kasus DBD tertinggi berada di Jawa Timur dan paling banyak terjadi di Kabupaten Malang. Pada tahun 2017 terdapat 451 kasus DBD yang menyebabkan 7 orang meninggal dunia. Kasus DBD semakin meningkat yaitu pada bulan Desember 2018 terdapat 681 orang yang menderita penyakit DBD sedangkan 3 orang meninggal dunia [4]. Penderita penyakit DBD ini perlu adanya suatu tindakan pencegahan maupun penanggulangan agar tidak mengalami peningkatan. Bentuk tindakan tersebut berupa melakukan peramalan. Hasil dari peramalan yang akurat dapat digunakan untuk alat pengambilan keputusan dalam proses pencegahan maupun penanggulangan penyakit DBD di Kabupaten Malang.

Penelitian tugas akhir ini mengacu pada penelitian sebelumnya yaitu tentang kasus jumlah kejadian demam berdarah dengan membandingkan beberapa metode termasuk metode *Random Forest Regression*.Penelitian tersebut diantaranya tentang perbandingan metode *Autoregressive Integrated Moving Average* (ARIMA), *Generalized Linear Autoregressive Moving Average* (GLARMA), dan model regresi *time series Random Forest* (RF) untuk melakukan prediksi jumlah Virus influenza A pada babi di Ontario, Kanada. Hasil peramalan meggunakan metode *Random Forest* memiliki nilai *error* yang kecil dibandingkan dengan yang lain [5]. Penelitian selanjutnya yaitu perbandingan metode *ARIMA* dan *Random Forest time series* dalam melakukan prediksi wabah avian influenza H5N1. Hasil dari penelitian menunjukkan bahwa model *Random Forest* lebih efektif untuk melakukan prediksi wabah avian influenza H5N1 [6].

Hasil beberapa penelitian tersebut menjadi dasar penggunaan metode *Random Forest Regression* dalam melakukan penelitian tugas akhir untuk peramalan jumlah KDB di Kabupaten Malang. Penggunaan metode *Random Forest Regression* ini masih belum diterapkan di Indonesia dalam kasus peramalan DBD di Kabupaten Malang.Oleh karena itu, penelitian tugas akhir ini mengusulkan peramalan jumlah KDB di Kabupaten Malang menggunakan metode *Random Forest Regression*. Selain itu, pada penelitian tugas akhir ini menggunakan variabel lag dari jumlah KDB. Hasil peramalan untuk periode yang mendatang tersebut diharapkan dapat membantu Dinas Kesehatan Kabupaten Malang dalam menekan jumlah KDB di Kabupaten Malang, Jawa Timur.

1.2. Rumusan Masalah

Berdasarkan latar belakang tersebut, rumusan masalah dalam studi kasus ini berfokus pada:

1. Bagaimana model terbaik yang dihasilkan dengan menggunakan metode *Random Forest*.
2. Bagaimana tingkat akurasi hasil peramalan yang didapatkan dari penerapan model terbaik metode *Random Forest*.
3. Apa saja informasi yang dapat diberikan kepada pihak Dinas Kesehatan Kabupaten Malang terkait hasil peramalan jumlah KDB.

1.3. Batasan Permasalahan

Berdasarkan rumusan permasalahan tersebut maka terdapat batasan dalam melakukan penelitian tugas akhir ini diantaranya:

1. Penelitian ini menggunakan *independent variable* yaitu suhu udara, curah hujan, kelembapan udara, kecepatan angin, angka bebas jentik dan jumlah penduduk di Kabupaten Malang.
2. Penelitian menggunakan data jumlah KDBsebagai *dependent variable* di Kabupaten Malang, Jawa Timur

dengan data bulanan mulai dari Januari 2010 sampai dengan Desember 2018.

3. Data yang digunakan adalah data jumlah kasus demam berdarah, kepadatan penduduk, dan Angka Bebas Jentik (ABJ) sebagai data sekunder yang didapatkan dari Dinas Kesehatan Kabupaten Malang. Sedangkan data iklim seperti curah hujan, kecepatan angin, kelembapan udara, dan suhu udara didapatkan dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Karangploso dan Karangkates.
4. Peramalan dilakukan untuk meramalkan jumlah KDBdi Kabupaten Malang dalam periode bulanan mulai Januari 2019 sampai dengan Desember 2020.
5. Peramalan menggunakan metode *Random Forest Regression.*
6. Penelitian dilakukan dengan menggunakan bahasa pemrograman *Python.*

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah maka terdapat tujuan dilakukan penelitian tugas akhir ini diantaranya:

1. Mendapatkan model terbaik yang dihasilkan dari penerapan metode *Random Forest.*
2. Mengetahui tingkat akurasi hasil peramalan dari penerapan model terbaik metode *Random Forest*.
3. Memberikan informasi terkait hasil peramalan jumlah KDB kepada Dinas Kesehatan Kabupaten Malang.

1.5. Manfaat Penelitian

Berdasarkan permasalahan dan tujuan maka terdapat beberapa manfaat yang didapatkan dari adanya penelitian tugas akhir ini diantaranya:

1. Memberikan informasi kepada pihak Dinas Kesehatan Kabupaten Malang untuk merencanakan penanganan dan tindakan dalam menekan jumlah KDB di Kabupaten Malang, Jawa Timur.
2. Menambah ilmu pengetahuan dalam bidang *forecasting* dan *data mining* dengan menggunakan metode *Random Forest Regression* jika diterapkan dalam bidang kesehatan.

1.6. Relevansi

Penelitian dalam tugas akhir ini berkaitan dengan bidang kesehatan yaitu tentang penyakit demam berdarah *dengue*. Permasalahan ini banyak terjadi di negara Asia Tenggara. Negara Indonesia termasuk dalam delapan negara yang mengalami penyakit demam berdarah *dengue* tertinggi khususnya di Jawa Timur. Penyakit demam berdarah *dengue* banyak terjadi khususnya di Kabupaten Malang, Jawa Timur. Penerapan metode *Random Forest Regression* dalam peramalan perlu dilakukan untuk mengetahui jumlah kasus demam berdarah pada periode berikutnya. Hasil peramalan akan membantu Dinas Kesehatan Kabupaten Malang dalam menekan jumlah KDBdi Kabupaten Malang.

Topik peramalan dalam penelitian tugas akhir ini memiliki relevansi dengan salah satu Laboratorium di Departemen Sistem Informasi, Institut Teknologi Sepuluh Nopember yaitu Laboratorium Rekayasa Data dan Intelegensi Bisnis (RDIB). Pada Tabel 1.1 menunjukkan relevansi roadmap antara penelitian tugas akhir dan laboratorium. Relevansi ini terdapat pada bidang *Business Intelligence* yaitu tentang *Business Analytics* dan *Data Mining*. Selain itu, penelitian tugas akhir ini juga memiliki relevansi dengan mata kuliah di Departemen Sistem Informasi tentang peramalan yang dipelajari di mata kuliah Teknik Peramalan dan penerapan metode *machine learning* di mata kuliah Penggalian Data dan Analitika Bisnis.

1.7. Sistematika Penulisan

Sistematika penulisan laporan tugas akhir disesuaikan dengan format yang telah ditentukan yaitu terdiri dari tujuh bab sebagai berikut.

BAB I PENDAHULUAN

Bab ini menjelaskan tentang latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, dan relevansi pengerjaan penelitian tugas akhir

BAB II TINJAUAN PUSTAKA

Pada bab Bab ini menjelaskan tentang penelitian sebelumnya dan dasar teori yang digunakan sebagai dasar pengetahuan maupun konsep dalam pengerjaan tugas akhir. Dasar teori ini mencakup berbagai teori dan informasi yang mendukung untuk menyelesaikan permasalahan dalam studi kasus tugas akhir.

Tabel 1.1. Roadmap Penelitian dan Laboratorium

Computerized Decision Support
• *Decision Support and Decision Support System* • *System Modeling and Analysis*
Business Intelligence
• *Data Warehousing* • *Business Analytics* • *Data IVsualization* • *Data, Text, and Web Mining* • *Business Performance Management*
CCGS and KM
• *Collaborative Computer Supported Technology* • *Group Support System* • *Knowledge Management*
Intelligent System
• *Artificial Intelligence* • *Expert System* • *Advance Intelligent System* • *Intelligent System over the Internet*

BAB III METODOLOGI

Bab ini menjelaskan tentang tahapan yang dilakukan dalam pengerjaan tugas akhir. Tahapan tersebut mencakup studi literatur, tahapan metode peramalan, evaluasi hasil dan kesimpulan dari peramalan.

BAB IV PERANCANGAN

Bab ini menjelaskan tentang perancangan yang lebih detail dan spesifik dalam membuat model terbaik untuk peramalan. Terdapat proses pengumpulan data, gambaran data sebagai *input* dan *output*, serta proses dalam pengolahan data yang digunakan untuk proses peramalan.

BAB V IMPLEMENTASI

Bab ini menjelaskan tentang proses penerapan metode *Random Forest Regression* dalam melakukan peramalan jumlah kasus demam berdarah.

BAB VI HASIL DAN PEMBAHASAN

Bab ini menjelaskan hasil dan pembahasan dari model yang telah didapatkan dengan menggunakan metode *Random Forest Regression*. Selain itu, terdapat hasil peramalan dari model tersebut untuk periode yang akan datang.

BAB VII KESIMPULAN DAN SARAN

Bab ini menjelaskan tentang kesimpulan yang didapatkan dari penerapan metode *Random Forest Regression*. Terdapat beberapa saran yang dapat dilakukan dalam pengembangan model sehingga penggabungan metode dapat lebih baik lagi.

Halaman ini sengaja dikosongkan

BAB II
TINJAUAN PUSTAKA

Pada bab ini berisi penelitian sebelumnya yang dijadikan acuan dalam pengerjaan tugas akhir dan juga berisi dasar teori untuk menunjang penelitian pada tugas akhir.

2.1. Penelitian Sebelumnya

Bagian ini menjelaskan beberapa penelitian sebelumnya yang digunakan sebagai dasar penelitian tugas akhir disajikan dalam Tabel 2.1.

Tabel 2.1. Studi Sebelumnya

PENELITIAN 1	
Judul	Assesment of Autoregressive Integreted Moving Average (ARIMA), Generalized Linear Autoregressive Moving Average (GLARMA), and Random Forest (RF) Time Series Regression Models for Predicting Influenza A Virus Frequency in Swine in Ontorio, Canada [5]
Nama Peneliti	Tatiana Petukhova, Davor Ojkic, Beverly McEwen, Rob Deardon, dan Zvonimir Poljak
Tahun Penelitian	2018
Deskripsi	Penelitian ini membahas tentang perbandingan metode *Autoregressive Integrated Moving Average* (ARIMA), *Generalized Linear Autoregressive Moving Average* (GLARMA), dan model regresi *time series Random Forest* (RF) untuk melakukan prediksi jumlah Virus influenza A pada babi di Ontario, Kanada. Data yang digunakan yaitu data diagnostik Virus influenza A tahun 2007 sampai tahun 2015. Data tersebut didapatkan dari the Animal Health Laboratory (University of Guelph, Guelph, ON, Canada). Pengukuran tingkat akurasi menggunakan *Root Mean Square Error* (RMSE) dan *Normalized Root Mean Square Error* (NRMSE).

Kelebihan	Penerapan metode *Random Forest* (RF) menghasilkan nilai *error* yang lebih kecil dibandingkan dengan metode *Autoregressive Integrated Moving Average* (ARIMA) dan *Generalized Linear Autoregressive Moving Average* (GLARMA). Hasil tersebut melibatkan faktor musim dan pengajuan virologi dalam periode bulanan dan mingguan.
Kekurangan	Penelitian ini berfokus hanya pada variabel musim dan pengajuan virologi. Variabel lain yang tidak menjadi pertimbangan seperti suhu udara dan kelembapan yang mungkin berpengaruh terhadap jumlah Virus influenza A.
Keterkaitan dengan Tugas Akhir	Penelitian ini mendukung untuk menerapkan metode *Random Forest* pada penelitian tugas akhir karena memiliki nilai *error* yang paling kecil. Perbedaan dengan penelitian tugas akhir yaitu studi kasus yang digunakan dalam penelitian.
PENELITIAN 2	
Judul	Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks [6]
Nama Peneliti	Michael J Kane, Natalie Price, Matthew Scotch, dan Peter Rabinowitz
Tahun Penelitian	2014
Deskripsi	Penelitian ini membahas tentang pemodelan *time series Random Forest* yang dapat memberikan peningkatan kemampuan dalam melakukan prediksi untuk wabah penyakit yang menular. Hasil dari penelitian memberikan pendekatan baru dalam memprediksi wabah berbahaya dan populasi burung berdasarkan pada data yang tersedia secara online. Hasil penelitian menunjukkan bahwa model *time series Random Forest* memiliki nilai *Mean Absolute*

	Percentage Error (MAPE) yang lebih kecil dibandingkan ARIMA.
Kelebihan	Penggunaan nilai *lagged* dari *variable importance* dapat meningkatkan akurasi metode *Random Forest* pada peramalan *time series*. Penelitian ini menggunakan nilai *lagged* 3 minggu pada *variable importance* dalam memprediksi wabah flu burung.
Kekurangan	Hubungan keterkaitan antara wabah yang diprediksi dengan penggunaan nilai *lagged* dari wabah bersifat tidak linear. Model ARIMA tidak mampu menangkap variabel yang tidak linear sehingga penggunaan nilai *lagged* pada penelitian ini mengakibatkan ARIMA memiliki kinerja yang buruk.
Keterkaitan dengan Tugas Akhir	Penggunaan metode *Random Forest* (RF) dalam bidang kesehatan. Perbedaan dengan penelitian tugas akhir yaitu studi kasus yang digunakan pada tugas akhir tentang peramalan jumlah kasus DBD.
PENELITIAN 3	
Judul	Relative Evaluation of Regression Tools for Urban Area Electrical Energy Demand Forecasting [7]
Nama Peneliti	Nils Jakob Johannesen, Mohan Kolhe, dan Morten Goodwin
Tahun Penelitian	2019
Deskripsi	Penelitian ini membahas tentang peramalan permintaan energi listrik di area perkotaan dengan melakukan evaluasi relative terhadap metode regresi. Evaluasi ini bertujuan menemukan metode yang cocok untuk peramalan tersebut dengan melibatkan beberapa metode regresi yaitu *Random Forest Regressor*, *k-Nearest Neighbour Regressor*, and *Linear Regressor*. Peramalan permintaan energi listrik di area perkotaan dilakukan untuk jangka waktu pendek (30 menit) dan jangka panjang (24 jam). Parameter yang digunakan dalam penelitian yaitu waktu, cuaca, dan *random effects.* Nilai *error*

	dari peramalan akan dihitung dengan menggunakan *Mean Absolute Percentage Error* (MAPE).
Kelebihan	Peramalan permintaan energi listrik menggunakan beberapa metode regresi dengan melalui pendekatan waktu vertikal (periode musiman dan mingguan). Selain itu, penelitian ini juga mempertimbangkan pengaruh dari parameter meteorologi. Hasil dari peramalan menunjukkan bahwa peramalan untuk jangka pendek lebih baik menggunakan metode *Random Forest Regressor* sedangkan jangka panjang menggunakan metode *k-Nearest Neighbour Regressor*.
Kekurangan	Pada proses analisis dampak dari *q-value* dalam perhitungan menggunakan *Mean Absolute Percentage Error* (MAPE) hanya melihat pada pola beban sebelumnya.
Keterkaitan dengan Tugas Akhir	Penelitian ini menggunakan metode *Random Forest* yang berfokus pada regresi. Selain itu, penelitian ini memberikan informasi bahwa *Random Forest Regressor* memiliki nilai *error* yang kecil dalam peramalan jangka pendek. Hal ini dapat menjadi pertimbangan dalam penelitian tugas akhir. Perbedaan dengan penelitian tugas akhir yaitu studi kasus dimana metode diterapkan pada bidang kesehatan.
PENELITIAN 4	
Judul	Machine Learning Methods Reveal the Temporal Pattern of Dengue Incidence Using Meteorological Factors in Metropolitan Manila, Philipines [8]
Nama Peneliti	Thaddeus M.Carvajal, Katherine M. Viacrusis, Lara Fides T.Hernandez, Howell T.Ho, DiIVna M. Amalin, dan Kozo Watanabe
Tahun Penelitian	2018

Deskripsi	Penelitian ini membahas tentang perbandingan beberapa metode *machine learning* untuk memprediksi pola temporal dari kejadian demam berdarah di Manila, Filipina. Dalam melakukan perbandingan, penelitian ini menggunakan 4 metode *machine learning* yaitu *General Additive Modelling, Seasonal Autoregressive Integrated Moving Average with exogenous variables, Random Forest, dan Gradient Boosting.* Peramalan tersebut menggunakan data kejadian demam berdarah dan faktor meteorologi diantaranya banjir, curah hujan, suhu udara, indeks osilasi selatan, kelembapan relatif, kecepatan, dan arah angin. Semua data tersebut didapatkan dari masing-masing lembaga pemerintah di Metropolitan Manila dari 1 Januari 2009 sampai dengan 31 Desember 2013.
Kelebihan	Pada penelitian ini menggunakan efek tertunda atau *lagged effects* (LG) pada faktor meteorologi dimana hasilnya lebih tepat dalam melakukan prediksi demam berdarah. Jeda waktu ini menyebabkan keterlambatan potensial pada waktu dimana cuaca memengerahui *vector* nyamuk dan Virus. Hasil peramalan menunjukkan bahwa metode *Random Forest with delayed meteorological effects* (RF-LG) memiliki nilai *error* yang kecil dibandingkan dengan metode yang lain.
Kekurangan	Penelitian yang dilakukan hanya untuk meramalkan kejadian demam berdarah di satu lokasi saja di Filipina yaitu Metropolitan, Manila. Hasil peramalan mungkin berbeda jika diterapkan pada daerah lain di Filipina. Selain itu, penelitian ini hanya mengambil faktor ekologis namun tidak mempertimbangkan faktor lain seperti aspek biologis dan sosiologis.

Keterkaitan dengan Tugas Akhir	Penelitian ini mendukung penerapan metode *Random Forest* (RF) dalam penelitian tugas akhir dengan studi kasus tentang penyakit demam berdarah *dengue*. Metode *Random Forest* memiliki nilai *error* yang lebih kecil. Perbedaan dengan penelitian tugas akhir yaitu lokasi studi kasus dan penggunaan variabel yang berpengaruh terhadap kejadian demam berdarah *dengue.*
PENELITIAN 5	
Judul	Forecasting the Number of Dengue Fever Cases in Malang Regency Indonesia Using Fuzzy Inference System Models [9]
Nama Peneliti	Wiwik Anggraeni, I Putu Agus Aditya Pramana, Febriliyan Samopa, Edwin Riksakomara, Radityo P. Wibowo, Lulus Condro T., dan Pujiadi
Tahun Penelitian	2017
Deskripsi	Penelitian ini membahas tentang peramalan jumlah kasus Demam Berdarah *Dengue* (DBD) di Kabupaten Malang. Peramalan menggunakan metode *Fuzzy Inference System* (FIS). Pengelompokan dalam pembentukan model berdasarkan lokasi geografis yaitu dataran tinggi, dataran sedang, dan dataran rendah. Hasil peramalan menunjukkan nilai *Mean Absolute Percentage Error* (MAPE) yang kecil di 3 dataran tersebut.
Kelebihan	Penerapan metode FIS dalam meramalkan jumlah kasus Demam Berdarah *Dengue* (DBD) cenderung memiliki nilai *error* kecil dan tingkat akurasi yang tinggi karena metode ini memerhatikan secara detail pengaruh dari setiap variabel. Selain itu, metode *Fuzzy Inference System* (FIS) tidak membutuhkan banyak data dan periode waktu yang terlalu panjang.

Kekurangan	Pada pengelompokan lokasi geografis masih belum detail dimana setiap daerah memiliki karakteristik yang berbeda-beda. Pertumbuhan kasus Demam Berdarah *Dengue* (DBD) dapat mempertimbangkan karakteristik dari masing-masing daerah di setiap dataran.
Keterkaitan dengan Tugas Akhir	Penelitian ini menggunakan studi kasus yang sama dengan penelitian tugas akhir. Perbedaan dengan penelitian tugas akhir yaitu penerapan metode yang digunakan untuk peramalan jumlah kasus Demam Berdarah *Dengue* (DBD).

2.2. Landasan Teori

Landasan teori berisi berbagai penjelasan teori yang dapat mendukung dalam pengerjaan penelitian tugas akhir.

2.2.1. Penyakit Demam Berdarah Dengue (DBD)

Penyakit Demam Berdarah Dengue (DBD) merupakan suatu penyakit demam tinggi yang biasanya banyak terjadi di daerah tropis dan subtropis. Penyakit DBD disebabkan adanya penyebaran Virus oleh nyamuk *Aedes aegypti*. Faktor penyebab kejadian DBD yaitu perubahan suhu udara, kelembapan, curah hujan, kondisi lingkungan, dan mobilitas penduduk. Penularan Virus nyamuk *Aedes aegypti* paling besar dipengaruhi oleh adanya mobilitas penduduk dimana penularan akan mengikuti pertumbuhan kepadatan penduduk. Tanda gejala penyakit DBD seperti demam yang tinggi disertai dengan pusing, adanya ruam dengan warna merah terang menyebar hampir ke seluruh tubuh, sakit pada persendian dan otot [2].

Berdasarkan data dari Kementrian Kesehatan (Kemenkes) RI mencatat bahwa penyebaran penyakit DBD tertinggi berada di Jawa Timur dengan jumlah kasus sebesar 700 orang sedangkan Jawa Tengah terdapat 512 orang, dan Jawa Barat terdapat 401 orang yang menderita penyakit tersebut. Penyebaran tersebut dimulai dari awal tahun 2018 hingga awal tahun 2019. Pihak Kementerian Kesehatan (Kemenkes) RI mengirimkan surat

edaran kepada kepala dinas kesehatan provinsi di seluruh Indonesia dengan tujuan menghimbau agar meningkatkan sosialiasi kepada masyarakat dan memberikan edukasi pemberantasan sarang nyamuk melalui kegiatan menguras, menutup, dan mendaur ulang barang bekas. Bentuk pencegahan lainnya yaitu dengan melalui pemantauan jumlah jentik secara berkala, dan memakai bahan kimia atau insektisida serta larvasida dalam memberantas nyamuk [10].

2.2.2. Peramalan Time Series

Time series adalah suatu rangkaian pengamatan terhadap data yang berdasarkan urutan waktu. Pada umumnya, *record* dari data time series memiliki interval yang konstan. Peramalan *time series* digunakan untuk meramalkan nilai data di masa mendatang dengan menggunakan data sebelumnya. Peramalan *time series* banyak digunakan dalam bisnis dan organisasi untuk melakukan perencanaan pada masa mendatang [11].

2.2.3. Uji Korelasi

Uji korelasi bertujuan untuk melihat hubungan keterikatan antara dua variabel pada data. Korelasi banyak digunakan pada tahapan penelitian termasuk dalam regresi. Salah satu jenis korelasi yang digunakan untuk mengukur tingkat akurasi antar variabel dengan nilai numerik yaitu *Pearson Correlation Coefficient (r)* dengan rentang nilai -1 sampai +1. Jika perhitungan korelasi mendekati + 1 maka antar variabel memiliki hubungan yang positif dan saling berpengaruh dimana saat variabel x meningkat maka variabel y juga meningkat. Namun, jika mendekati -1 maka suatu variabel x bertambah maka vaiabel y akan berkurang. Hal ini karena hasilnya mendekati -1. Rumus koefisien korelasi Pearson ditunjukkan pada persamaan (2-1) [7] .

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{\{n\sum_x 2 - (\sum_x)2\}\{n\sum_y 2 - (\sum_y)2\}}} \qquad (2\text{-}1)$$

dimana

n = jumlah pasangan data X dan Y

x = jumlah dari variabel x
y = jumlah daru variabel y
xy = hasil perkalian dari jumlah variabel x dany

2.2.4. *Pra-processing Data*

Pra-processing data merupakan proses yang dilakukan pengelolaan data sebelum digunakan untuk membentuk suatu model peramalan. Proses ini dilakukan dengan 2 langkah yaitu *missing value* dan normalisasi.

1. *Missing Value*

Prosedur *imputation* merupakan prosedur yang menangani permasalahan *missing value* dengan cara menggantikan beberapa nilai tertentu. *Mean computation* merupakan salah satu cara menangani *missing value* yang terkenal mudah. Penanganan *missing value* juga dapat dengan cara mengganti masing-masing *missing value* dengan metode interpolasi menggunakan *window interpolation* berdasarkan variabel pada data *time series* [12].

2. Normalisasi

Normalisasi adalah suatu bentuk tranformasi *"scalling down"* untuk menghindari perbedaan yang besar diantara *variable input*. Proses ini dilakukan pada data *training* maupun data *testing*. Normalisasi yang sering digunakan yaitu normalisasi min-max dalam skala interval (0,1). Rumus perhitungan normalisasi min-max ditunjukkan pada persamaan (2-2) [13].

$$y_i' = \frac{y_{max} - y_i}{y_{max} - y_{min}} \qquad (2\text{-}2)$$

dimana
y_i' = nilai skala

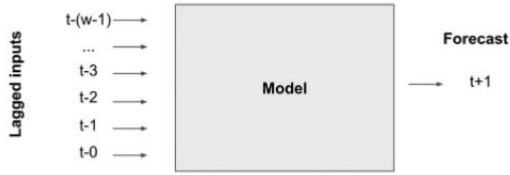
y_i = nilai asli atau aktual

2.2.5. Windowing

Data *time series* merupakan data yang unik dengan berisi informasi untuk digunakan dalam memprediksi data di masa depan dengan menggunakan data sebelumnya. Kumpulan dari

data sebelumnya sebagai data *input* ke dalam model sehingga dapat digunakan untuk menghitung nilai data *output* yang menjadi target di masa depan. Teknik standard yang digunakan dalam *machine learning* yaitu *windowing*. Teknik ini digunakan untuk membangun model berdasarkan hubungan antara input dengan target.

Data *time series* diubah menjadi data *cross-sectional*. Pada penerapan *windowing* terdapat satu data *set time series* yang berurutan yang dijadikan sebagai *window*. *Record* terakhir dari data *set time series* akan menjadi target yang dihitung dari *lagged variable input.* Pada gambar 2.1 dijelaskan hubungan antara *lagged inputs* jika dijadikan model untuk menghasilkan *output* target yaitu peramalan di masa mendatang [11].

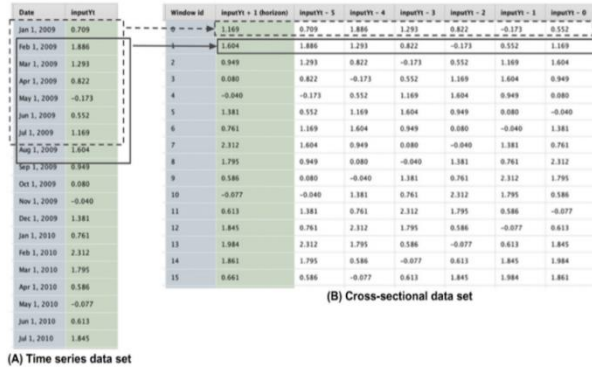


Gambar 2.1. Model Machine Learning pada Time Series

Jumlah *window* yang cukup diektraksi dari data *set time series* dimana model *supervised* dapat dipelajari berdasarkan hubungan antara *lagged variable input* dan variabel target. Nilai dari variabel target yang dihasilkan dapat digunakan untuk menentukan jendela baru dan melakukan prediksi untuk masa mendatang. Proses ini dapat dilakukan berulang kali sampai semua prediksi selesai dibuat. Tujuan teknik *windowing* ini yaitu untuk mengubah data *time series* menjadi data *set input machine learning*. Pada gambar 2.2. ditunjukkan proses perubahan dari data *time series* menjadi data *set cross-sectional* [11].

Pada Gambar 2.2. menunjukkan bahwa contoh tersebut menggunakan *window size* = 6, *step* = 1, *horizon width* = 1, dan *skip* = 0. Dalam proses *windowing* dan *cross-sectional* dapat dilakukan perubahan parameter *window, overlap* antara *window*

yang berurutan, dan prediksi horizon yang berperan sebagai peramalan.



Date	inputYt
Jan 1, 2009	0.709
Feb 1, 2009	1.886
Mar 1, 2009	1.293
Apr 1, 2009	0.822
May 1, 2009	-0.173
Jun 1, 2009	0.552
Jul 1, 2009	1.169
Aug 1, 2009	1.604
Sep 1, 2009	0.949
Oct 1, 2009	0.080
Nov 1, 2009	-0.040
Dec 1, 2009	1.381
Jan 1, 2010	0.761
Feb 1, 2010	2.312
Mar 1, 2010	1.795
Apr 1, 2010	0.586
May 1, 2010	-0.077
Jun 1, 2010	0.613
Jul 1, 2010	1.845

(A) Time series data set

Window id	inputYt + 1 (horizon)	inputYt - 5	inputYt - 4	inputYt - 3	inputYt - 2	inputYt - 1	inputYt - 0
0	1.169	0.709	1.886	1.293	0.822	-0.173	0.552
1	1.604	1.886	1.293	0.822	-0.173	0.552	1.169
2	0.949	1.293	0.822	-0.173	0.552	1.169	1.604
3	0.080	0.822	-0.173	0.552	1.169	1.604	0.949
4	-0.040	-0.173	0.552	1.169	1.604	0.949	0.080
5	1.381	0.552	1.169	1.604	0.949	0.080	-0.040
6	0.761	1.169	1.604	0.949	0.080	-0.040	1.381
7	2.312	1.604	0.949	0.080	-0.040	1.381	0.761
8	1.795	0.949	0.080	-0.040	1.381	0.761	2.312
9	0.586	0.080	-0.040	1.381	0.761	2.312	1.795
10	-0.077	-0.040	1.381	0.761	2.312	1.795	0.586
11	0.613	1.381	0.761	2.312	1.795	0.586	-0.077
12	1.845	0.761	2.312	1.795	0.586	-0.077	0.613
13	1.984	2.312	1.795	0.586	-0.077	0.613	1.845
14	1.861	1.795	0.586	-0.077	0.613	1.845	1.984
15	0.661	0.586	-0.077	0.613	1.845	1.984	1.861

(B) Cross-sectional data set

Gambar 2.2. Proses Windowing

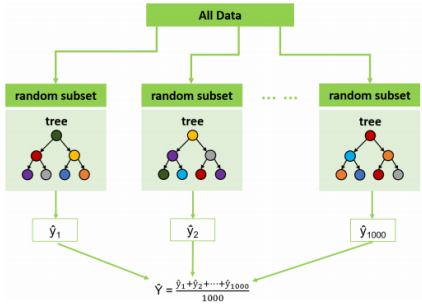
Parameter dalam melakukan *windowing* [11]:

1. *Window Size*: Jumlah titik *lag* dalam satu *window* namun tidak termasuk target.
2. *Step*: Jumlah titik data antara nilai pertama dari dua *window* berturut-turut.
3. *Horizon width*: Jumlah dari prediksi horizon sebagai variabel target di masa mendatang.
4. *Skip: Offset* antara *window* dengan prediksi horizon.

2.2.6. Random Forest Regression

Machine learning terbagi menjadi dua kategori yaitu kategori *supervised* dan *unsupervised*. *Supervised* merupakan data yang mempunyai nilai yang diketahui sedangkan *unsupervised* berarti nilai atau hasil dari suatu data yang tidak diketahui. Kategori *supervised* terbagi menjadi dua jenis yaitu regresi dan klasifikasi. Pada penerapannya jika hasil yang diperoleh berupa *continous* maka dapat dikatakan sebagai model regresi. Namun, jika yang dihasilkan berupa kategorikal maka termasuk dalam model klasifikasi. *Random Forest Regression* merupakan metode *machine learning* kategori *supervised* dengan hasil yang diperoleh berupa numerik [8].

Random Forest merupakan salah satu pengembangan dari metode *Decision Tree*. Hal ini karena terdapat pola perulangan beberapa kali terhadap penerapan metode *Decision Tree.* Metode ini menggabungkan hasil dari *Decision Tree* untuk diambil nilai rata-rata sehingga menghasilkan nilai akhir dengan tingkat akurasi yang lebih baik. Penerapan metode *Random Forest* menggunakan agregasi *bootstrap (bagging)* untuk mengukur akurasi dari estimasi sampel. *Sample bootstrap* diambil dari kumpulan data training dimana akan digunakan untuk membentuk *Decision Tree* [8]. Pada gambar 2.3. menjelaskan tentang rumus perhitungan dari *Random Forest.* Hasil dari prediksi didapatkan dari perhitungan rata-rata dari semua prediksi *Decision Tree* yang telah dilakukan. Gambar 2.3. merupakan gambar dari *framework* algoritma *Random Forest* [14].



Gambar 2.3. Framework Algoritma Random Forest

Jadi rumus dari metode *Random Forest* secara umum ditunjukkan pada persamaan (2-3) [14].

$$\hat{Y} = \frac{\hat{Y}_1 + \hat{Y}_2 + \hat{Y}_3 + \ldots \hat{Y}_N}{N} \qquad (2\text{-}3)$$

Pada Gambar 2.3. terlihat bahwa terdapat N sebanyak1000 diambil dari data training dimana digunakan untuk membentuk setiap *node* dari *decision tree*. Pada setiap *node* pemilihan dilakukan secara acak untuk mengoptimalkan pemisahan sehingga hasil prediksi diperoleh dari rata-rata prediksi dari semua *tree* [14].

2.2.7. Evaluasi Hasil Peramalan

Evaluasi hasil peramalan perlu dilakukan untuk menghasilkan peramalan yang akurat. Pada penelitian Tugas Akhir ini menggunakan RMSE dan sMAPE.

1. *Root Mean Square Error* (RMSE)

   RMSE dihitung dengan cara menghitung nilai rata-rata dari jumlah kuadrat *error*. RMSE pada umumnya digunakan untuk mengevaluasi kinerja suatu model dan menentukan akurasi analisis tipe regresi dengan hasil variabel *continuous (numerical*)[8]. Sedangkan MSE lebih rentan terhadap outlier karena error pada outlier akan diberikan bobot lebih besar. Hal ini akan membuat nilai MSE semakin besar. RMSE merupakan akar kuadrat dari MSE dan berada pada skala yang sama dengan data yang sedang dievaluasi[15]. RMSE lebih tepat digunakan saat model terhadap kesalahan mengikuti distribusi normal [16]. Rumus perhitungan RMSE ditunjukkan pada persamaan (2-7) [17].

$$\text{RMSE} = \sqrt{\frac{1}{T_f} \sum_{i=1}^{T_f} (y_i' - y_i)^2} \qquad (2\text{-}7)$$

2. *symmetric Mean Absolute Percentage Error* (sMAPE)

   sMAPE digunakan sebagai pengujian dalam melakukan perbandingan tingkat akurasi model antara hasil peramalan dengan data aktual. sMAPE berfungsi untuk menghindari masalah interpretasi yang terjadi ketika nilai aktual ada yang bernilai nol. Sedangkan *Mean Absolute Percentage Error* (MAPE) tidak dapat digunakan jika terdapat nilai aktual yang bernilai nol [15] [18]. Rumus perhitugan *symmetric Mean Absolute Percentage Error* ditunjukkan pada persamaan (2-8) [19].

$$\text{sMAPE} = \frac{2}{k} \sum_{i=1}^{k} \frac{| Yt - |\widehat{\text{Yt}}}{| Yt | + | \widehat{\text{Yt}} |)} * 100\% \qquad (2\text{-}8)$$

BAB III
METODOLOGI PENELITIAN

Bab Metodologi Penelitian akan menjelaskan bagaimana alur dari penelitian beserta penjelasan dari tiap tahapannya sebagai metodologi yang digunakan.

3.1. Metodologi

Pada Gambar 3.1. menjelaskan tentang diagram alur dari tahapan pengerjaan tugas akhir.

3.2. Uraian Metodologi

Dibawah ini merupakan penjelasan lebih rinci tentang tahapan yang dilakukan dalam pengerjaan tugas akhir.

3.2.1. Studi Literatur

Studi literatur merupakan tahapan mengumpulkan berbagai informasi tentang peramalan data *time series* dan metode *Random Forest Regression*. Informasi didapatkan dari berbagai sumber seperti buku, jurnal, dan penelitian-penelitian terkait sebelumnya.

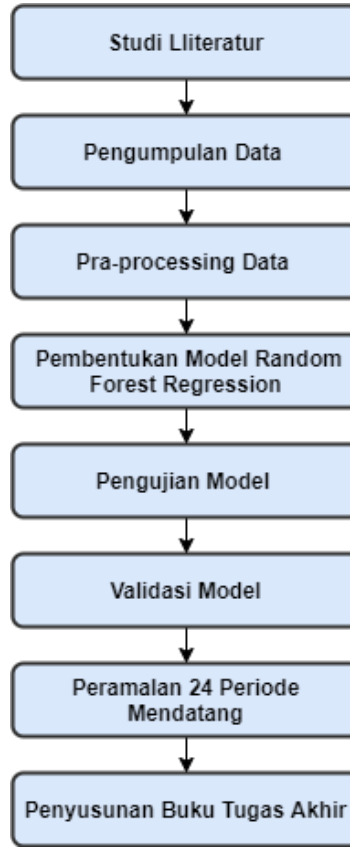
Tahapan awal dengan memahami dan melakukan analisis terhadap permasalahan data jumlah KDBdi Kabupaten Malang. Selain itu, tahapan ini mempelajari *pra-processing*, algoritma *Random Forest Regression*, dan penerapan metode tersebut dalam peramalan *time series*.

Tujuan dari identifikasi masalah adalah menentukan batasan masalah, tujuan, manfaat, dan mendukung latar belakang terkait dengan penerapan metode *Random Forest Regression*.

3.2.2. Pengumpulan Data

Tahap pengumpulan data dilakukan sebagai pendukung utama dalam proses penelitian tugas akhir. Data yang dibutuhkan dalam penelitian tugas akhir ini adalah data jumlah KDBdi Kabupaten Malang dari periode bulanan yaitu mulai dari bulan Januari 2010 sampai dengan bulan Desember 2018. Selain itu,

pengambilan data untuk curah hujan, kecepatan angin, kelembapan udara, dan suhu udara didapatkan dari website BMKG Karangploso dan Katangkates sebagai data tambahan dalam penelitian tugas akhir.



Gambar 3.1. Alur Pengerjaan Tugas Akhir

3.2.3. Pra-processing Data

Tahapan ini digunakan untuk pengelolaan data yang digunakan dalam penelitian tugas akhir. Terdapat 5 langkah dalam melakukan pra-processing yaitu missing value, uji korelasi variabel, normalisasi, pembentukan lag, dan pembagian data.

1. Missing Value

Penanganan *missing value* yang terdapat pada data dapat dilakukan dengan cara mengisi *missing value* dengan cara melakukan rata-rata, interpolasi dan regersi yang dapat merepresentasikan dari beberapa nilai variabel yang lain pada data jumlah KDB.

2. Uji Korelasi Variabel

Uji korelasi bertujuan untuk melihat keterkaitan atau hubungan antara *variable dependent* dan *variable independent*. Pada kasus ini *variable dependent* yaitu jumlah KDB sedangkan terdapat beberapa *independent variable* seperti suhu udara, temperature, angka bebas jentik, kepadatan penduduk, curah hujan, kecepatan angin, dan kelembapan udara.

Pada pengerjaan tugas akhir ini melakukan 2 proses uji korelasi yaitu uji korelasi terhadap pemilihan kecamatan dan pemilihan *dependent variable*. Pemilihan kecamatan hanya menggunakan jumlah KDB semua kecamatan di setiap dataran. Sedangkan, pemilihan *independent variable* menggunakan jumlah KDB dan semua *dependent variable.*

Hasil pengujian korelasi dapat digunakan untuk memilih *independent variable* mana yang memengaruhi jumlah KDB. Proses ini dilakukan sesuai formula *Pearson Correlation Coefficient (r)* dengan rentang nilai -1 sampai +1. *Independent variable* yang paling berpengaruh terhadap *dependent variable* maka variabel tersebut yang akan digunakan dalam proses pemodelan dan peramalan.

3. Normalisasi

Normalisasi dilakukan pada data training dan data testing yang bertujuan untuk menghindari perbedaan dari data yang bernilai besar diantara *variable input* lain. Proses normalilasi menggunakan min-max pada persamaan (2-2).

4. Pembentukan Lag

Proses pembentukan lag menggunakan teknik windowing. Teknik ini merupakan suatu teknik dalam pembangunan model berdasarkan hubungan antara input dengan target dimana data *time series* diubah menjadi data *cross-sectional*. Penentuan jumlah *lag* dapat ditentukan melalui percobaan.

5. Pembagian Data

Tahapan ini dilakukan pembagian data yaitu data *training* dan data *testing*. Data *training* digunakan untuk pembentukan beberapa model sedangkan data testing digunakan untuk proses validasi model yang dihasilkan dari data *training*.

Tujuan validasi model yaitu mendapatkan model terbaik dari beberapa model yang akan digunakan dalam proses peramalan untuk periode mendatang. Pada tahapan ini terdapat 4 skenario pembagian data yang dilakukan. Skenario pertama data dibagi menjadi 80% data *training* dan 20% data *testing*, skenario kedua data dibagi menjadi 70% data *training* dan 30% data *testing*, skenario ketiga data dibagi 60% data *training* dan 40% data *testing*, dan skenario keempat data dibagi menjadi 50% data *training* dan 50% data *testing*.

3.2.4. Pembentukan Model Random Forest

Pembentukan model dilakukan berbagai percobaan kombinasi parameter yang bertujuan untuk mendapatkan hasil model yang terbaik. Pemodelan menggunakan *dependent variable* yaitu jumlah KDB dan *independent variabel* yang dipilih pada saat pengujian korelasi. Dalam pembentukan model dilakukan penerapan 4 skenario pembagian data antara data *training* dan data *testing* yang telah ditentukan pada sub bab pembagian data.

Terdapat beberapa tahapan lebih rinci dalam mencari model terbaik dengan menggunakan metode *Random Forest Regression* yaitu tuning parameter dan skenario pembentukan model.

1. Tuning Parameter

Percobaan menggunakan berbagai parameter dari *Random Forest Regression* di *Python* untuk menemukan model terbaik. Percobaan nilai parameter dilakukan bertujuan untuk mengoptimalkan nilai akurasi. Parameter yang diubah dalam proses ini yaitu penentuan *n_estimator, max_depth,* dan *min_samples_split.*

2. Skenario Pembentukan Model

Dalam melakukan pembentukan model, pengerjaan tugas akhir ini menggunakan 5 skenario variabel percobaan yang di setiap skenario melibatkan variabel lag jumlah KDB. Skenario tersebut meliputi model yang melibatkan semua *independent variable* bedasarkan pada lolos uji korelasi dan lag KDB, 3 independent variable yang memiliki tingkat korelasi tertingi dan lag KDB, 2 independent variable yang memiliki tingkat korelasi tertinggi dan lag KDB, *1 independent variable* yang memiliki tingkat korelasi tertinggi dan lag KDB, serta model hanya menggunakan variabel lag dari jumlah KDB. Skenario pembentukan model ini diterapkan pada semua dataran.

3.2.5. Pengujian Model Random Forest

Pada tahap ini dilakukan pengujian model yang didapatkan dalam pembentukan model terhadap data *testing*. Hal ini bertujuan untuk menguji tingkat akurasi model dengan menggunakaan RMSE dan sMAPE. Namun, model terbaik dipilih berdasarkan pada nilai RMSE sebagai pengukuran tingkat akurasi primer bukan berdasarkan nilai sMAPE.

3.2.6. Validasi Model Random Forest

Proses validasi model melalui 2 tahapan yaitu validasi pembagian data dan kecamatan lain. Validasi dilakukan terhadap model terbaik yang telah didapatkan dari 5 skenario variabel.

3.2.7. Peramalan 24 Periode Mendatang

Proses peramalan dilakukan dengan menggunakan model terbaik yang telah didapatkan pada proses pengujian dan

validasi model Random Forest. Tahapan ini bertujuan untuk meramalkan jumlah KDB untuk periode bulanan mulai dari tahun 2019 sampai dengan tahun 2020.

3.2.8. Penyusunan Buku Tugas Akhir

Proses terakhir dari penelitian ini yaitu penuyusunan buku tugas akhir. Buku ini bertujuan untuk proses dokumentasi dalam penelitian tugas akhir yang berisi mulai dari langkah-langkah pengerjaan sampai dengan penarikan kesimpulan berdasarkan hasil yang telah didapatkan.

Halaman ini sengaja dikosongkan

BAB IV
PERANCANGAN

Pada bab ini dijelaskan mengenai perancangan dalam pembentukan model menggunakan algoritma *Random Forest Regression*. Perancangan ini meliputi pengumpulan data, *pra-processing* data, pembentukan dan pengujian model serta proses validasi model terbaik sebelum melakukan peramalan untuk 24 periode mendatang.

4.1. Pengumpulan Data

Data yang dibutuhkan dalam penelitian tugas akhir ini adalah data jumlah KDB di Kabupaten Malang dari periode bulanan yaitu mulai dari bulan Januari 2010 sampai dengan bulan Desember 2018 sebanyak 108 data. Pada data jumlah KDBterdapat beberapa independent variable yaitu angka bebas jentik, jumlah penduduk, curah hujan, kecepatan angin, kelembapan udara, dan suhu udara.

Pengambilan data untuk variabel iklim yaitu curah hujan, kecepatan angin, kelembapan udara, dan suhu udara didapatkan dari situs BMKG Karangploso dan Karangkates. Data angka bebas jentik, jumlah penduduk, dan jumlah kasus demam berdarah berupa data sekunder yang didapatkan dari Dinas Kesehatan Kabupaten Malang. Data berisi 16 kecamatan yang dibagi menjadi 3 kategori pengelompokan untuk data jumlah kasus demam berdarah. Kategori dikelompokkan berdasarkan letak geografis yaitu dataran tinggi berada pada ketinggian 500 – 700 mdpl, dataran sedang berada pada ketinggian 200 – 500 mdpl, dan dataran rendah berada pada ketinggian 0 – 200 mdpl.

4.2. *Pra-Processing* Data

Pada *pra-processing* data terdapat beberapa tahapan sebelum dilakukan proses pembentukan model. Proses ini perlu dilakukan beberapa langkah yaitu pengolahan data harian variabel iklim, *missing value*, uji korelasi data, normalisasi data, pembentukan lag dan pembagian data.

4.2.1 Pengolahan Data Harian Variabel Iklim

Data yang dibutuhkan dalam penelitian tugas akhir yaitu data bulanan dari periode 2010 sampai 2018 sedangkan data iklim yang dihasilkan dari situs BMKG berupa data harian. Dalam pengolahan data harian menjadi bulanan perlu dilakukan *pra-processing* data yang dilakukan dengan cara merata-rata data iklim harian di setiap bulan untuk seluruh data dari periode 2010 sampai 2018.

4.2.2 Missing Value

Tahapan ini digunakan untuk melakukan pengecekan dan pengisian *missing value* pada data sebelum dilakukan pembentukan model. Data yang didapatkan dari Dinas Kesehatan Kabupaten Malang banyak nilai yang kosong pada data Angka Bebas Jentik (ABJ). Data ABJ dibutuhkan sebagai *independent variable* dalam pembentukan model. Pengisian *missing value* terhadap data ABJ dilakukan dengan menggunakan 2 cara yaitu interpolasi dan regresi. Interpolasi digunakan untuk mengisi data yang tidak terlalu banyak *missing value* di setiap periode karena interpolasi tidak cukup baik diimplementasikan pada data yang banyak memiliki missing value [12]. Sedangkan regresi digunakan mengisi data pada periode yang memiliki banyak *missing value*.

Selain itu, terdapat *missing value* pada beberapa data iklim yang dihasilkan sehingga perlu melakukan pengisian *missing value*. Untuk mengisi *missing value* dilakukan dengan cara merata-rata nilai pada bulan yang sama di setiap periode.

4.2.3 Uji Korelasi Data

Proses uji koreasi memiliki rentang nilai -1 sampai +1. Jika nilai mendekati -1 menunjukkan hubungan yang berbanding terbalik sedangkan jika nilai mendekati +1 maka menunjukkan hubungan yang berbanding lurus antar variabel.

Proses uji korelasi dibedakan menjadi 2 tahapan yaitu uji korelasi terhadap pemilihan kecamatan dan pemilihan *independent variable*. Pemilihan kecamatan dan *independent*

variable didasarkan pada hasil uji korelasi yang memiliki nilai tinggi.

a. Pemilihan kecamatan

Pemilihan kecamatan dilakukan di setiap 3 kategori dataran yaitu dataran rendah, sedang, dan tinggi. Setiap dataran dipilih satu kecamatan yang akan digunakan untuk pembentukan model berdasarkan rata-rata hasil uji korelasi. Pengujian korelasi dilakukan antar jumlah kasus demam berdarah di kecamatan pada setiap dataran. Kecamatan yang memiliki rata-rata hasil uji korelasi tertinggi akan digunakan sebagai kecamatan pembentuk model.

b. Pemilihan independent variable

Pengujian korelasi dilakukan antara *independent variable* yaitu angka bebas jentik, jumlah penduduk, curah hujan, kecepatan angin, kelembapan udara, dan suhu udara terhadap jumlah kasus demam berdarah sebagai *dependent variable*.

4.2.4 Normalisasi Data

Proses normalisasi dilakukan terhadap semua data yaitu suhu udara, curah hujan, kecepatan angin, kelembapan udara, angka bebas jentik, jumlah penduduk, dan jumlah kasus DBD. *Range* nilai yang dimiliki pada data suhu udara, kecepatan angin, kelembapan udara, dan curah hujan berupa puluhan. Data angka bebas jentik memiliki nilai ratusan, data jumlah penduduk memiliki nilai ribuan sedangkan data jumlah KDB dan kecepatan angin memiliki *range* nilai satuan. Jumlah data terdiri dari 108 periode sehingga data yang dilakukan proses normalisasi sebanyak 756 data.

Proses ini perlu dilakukan karena nilai antar variabel memiliki *range* nilai yang tinggi. Normalisasi dilakukan dengan menggunakan fungsi min-max dengan skala interval 0 sampai dengan 1 sehingga antar variabel memiliki *range* nilai yang merata.

4.2.5 Pembentukan Lag

Penelitian tugas akhir ini menggunakan percobaan lag 0 sampai dengan lag 12. Variabel lag yang digunakan merupakan variabel jumlah KDB yang akan berperan sebagai variabel *input* atau *independent variable* dalam melakukan pembentukan model.

Pada penelitian tugas akhir ini, penambahan lag akan mengurangi periode yang digunakan pada setiap percobaan yang dilakukan. Jika percobaan awalnya menggunakan 6 kolom *independent variable* dan 1 kolom *dependent variable* yang masing-masing kolom memiliki 108 baris atau periode maka penerapan lag 12 akan memotong 12 periode awal pada data tersebut. Namun, kolom pada data akan bertambah juga sebanyak 12 kolom sehingga percobaan yang dilakukan melibatkan 18 kolom sebagai *independent variable* dan 1 kolom *dependent variable* dengan jumlah baris sebanyak 96.

4.2.6 Pembagian Data

Data yang terdiri dari *independent variable* dan *dependent variable* dibagi menjadi data pelatihan dan data pengujian. Pada pembentukan model terdapat skenario pembagian data. Tujuan dari skenario in yaitu untuk menentukan pengaruh dari *splitting* data dengan proporsi yang berbeda pada pembentukan model. Skenario pembagian data memengaruhi tingkat akurasi pada model sehingga performa tidak banyak mengalami penurunan (performa model tidak banyak terdegradasi) [20]. Pembagian data pada penelitian tugas akhir ini menggunakan 4 skenario seperti yang ditunjukkan pada Tabel 4.1.

Tabel 4.1. Skenario Pembagian Data

Skenario	Pembagian Data	
	Data Training	Data Testing
1	80%	20%
2	70%	30%
3	60%	40%
4	50%	50%

Data pelatihan digunakan dalam pembentukan model dengan metode *Random Forest Regression*. Untuk data pengujian digunakan pada proses validasi model yang dihasilkan dari *data training* dengan tujuan mendapatkan model terbaik dari beberapa model selama proses pembentukan model.

4.3. Pembentukan Model Random Forest

Pada proses pembentukan model terdapat 2 tahapan meliputi penentuan inisiasi parameter dan skenario dalam pembentukan model dengan menggunakan algoritma *Random Forest Regression.*

4.3.1 Penentuan Inisiasi Parameter

Proses *tuning parameter* digunakan untuk menemukan model terbaik dan mengoptimalkan nilai akurasi. Percobaan yang dilakukan pada penelitian tugas akhir ini yaitu mengubah nilai parameter *n_estimator, min_samples_split*, dan *max_depth* dari algoritma *Random Forest Regression*. Nilai dari setiap parameter yang akan dilakukan percobaan pada penelitian tugas akhir ini ditunjukkan pada Tabel 4.2.

Tabel 4.2. Nilai Inisiasi Parameter

Parameter	Nilai
Jumlah n_estimator	100 sampai 1000
Jumlah max_depth	5, 10, 15, 20
Jumlah min_samples_split	4, 8, 12, 16

a. Jumlah n_estimator

   Parameter *n_estimator* merupakan jumlah *tree* yang akan digunakan dalam membangun model *Random Forest Regression* [21]. Penerapan jumlah *tree* yang dilakukan pada tugas akhir ini melalui proses uji coba seperti yang ditunjukkan pada Tabel 4.2.

b. Jumlah max_depth

   Parameter *max_depth* adalah kedalaman setiap *tree* atau pohon yang akan dibangun. Parameter ini juga dapat diartikan seberapa tinggi pohon yang akan dibangun untuk

setiap melakukan percobaan. Semakin besar nilai *max_depth* maka akan semakin tinggi pula pohon yang akan dibangun [21].

c. Jumlah min_samples_split

Parameter *min_samples_split* merupakan jumlah simpul internal yang akan dibuat. Jika *max_depth* menentukan seberapa tinggi pohon tersebut maka min_samples_split menentukan seberapa banyak cabang besar yang akan dibuat. Semakin banyak jumlah parameter ini maka akan semakin banyak juga percabangan yang akan dibuat pada setiap *tree* [21].

4.3.2 Pembentukan Model Random Forest

Pada proses pembentukan model menggunakan metode *Grid Search* sehingga dapat membantu dalam menemukan model terbaik dari setiap skenario. Pembentukan model dilakukan pada kecamatan pembentuk model yang dipilih berdasarkan rata-rata hasil uji korelasi tertinggi di setiap dataran.

Kecamatan pembentuk model yang digunakan di setiap dataran diperoleh dari rata-rata hasil uji korelasi. Terdapat 3 kecamatan yang digunakan untuk pembentukan model sebagai perwakilan dari 3 dataran. Setiap pembentukan model menggunakan 4 skenario pembagian data, variabel lag 0 sampai lag 12, dan 5 skenario *independent variable*. Total percobaan yang dilakukan dengan melibatkan lag 0 sampai lag 12, skenario *independent variable* dan skenario pembagian data maka terdapat 260 kali percobaan.

a. Skenario tuning parameter

Skenario *tuning parameter* merupakan skenario yang dilakukan berbagai percobaan pada nilai parameter. Pada metode *Random Forest Regression* yang digunakan pada tugas akhir ini menggunakan 3 parameter yaitu n_estimator, min_samples_split, dan max_depth. Nilai parameter yang dilakukan percobaan seperti yang ditunjukkan pada Tabel 4.2. Jumlah percobaan pada setiap kali proses pembentukan model

sebanyak 160 kali percobaan untuk *tuning parameter* jika tidak menggunakan *Grid Search.*

b. Skenario independent variable

Skenario *independent variable* dilakukan pada setiap dataran dalam proses pembentukan model. Pada setiap melakukan percobaan terdapat skenario yang menggunakan lag dan tanpa lag serta terdapat 5 jenis skenario *independent variable* yang ditunjukkan pada Tabel 4.3.

Tabel 4.3. Skenario Independent Variable

Skenario	Keterangan
All Independent Variable (All IV)	Skenario yang melibatkan semua variabel yaitu suhu udara, curah hujan, kecepatan angin, kelembapan udara, angka bebas jentik, jumlah penduduk, dan variabel lag 0-12.
Top 1 Independent Variable (Top 1 IV)	Skenario yang melibatkan variabel suhu udara, curah hujan, kecepatan angin, kelembapan udara, angka bebas jentik, jumlah penduduk yang memiliki 1 korelasi tertinggi dan variabel lag 0-12.
Top 2 Independent Variable (Top 2 IV)	Skenario yang melibatkan variabel suhu udara, curah hujan, kecepatan angin, kelembapan udara, angka bebas jentik, jumlah penduduk yang memiliki 2 korelasi tertinggi dan variabel lag 0-12.
Top 3 Independent Variable (Top 3 IV)	Skenario yang melibatkan variabel suhu udara, curah hujan, kecepatan angin, kelembapan udara, angka bebas jentik, jumlah penduduk yang memiliki 3 korelasi tertinggi dan variabel lag 0-12.
No Independent Variable (No IV)	Skenario yang hanya melibatkan variabel lag 0-12 dari jumlah KDB.

c. Skenario Pembagian Data

Skenario pembagian data *testing* dan data *training* dilakukan pada setiap dataran dalam pembentukan model seperti yang ditunjukkan pada Tabel 4.1. Jumlah percobaan yang dilakukan pada setiap dataran sebanyak 4 kali *skenario independent variable*. Jika percobaan dilakukan pada skenario yang melibatkan semua *independent variable* maka sekali percobaan melakukan 48 kali percobaan di setiap dataran.

4.4. Pengujian Model Random Forest

Pemilihan model terbaik pada penelitian tugas akhir ini berdasarkan pada nilai RMSE yang terkecil. Model yang memiliki nilai RMSE paling kecil akan dipilih menjadi model terbaik dan digunakan untuk peramalan 24 periode mendatang. Meskipun model memiliki nilai SMAPE kecil namun nilai RMSE bukan yang terkecil maka tetap tidak akan dipilih sebagai model terbaik sehingga saat memilih model terbaik hanya fokus terhadap nilai RMSE saja Sedangkan SMAPE digunakan untuk mengetahui nilai akurasi model terbaik dan proses validasi model terbaik.

4.5. Validasi Model Random Forest

Proses validasi digunakan untuk melihat nilai akurasi dari model terbaik yang telah didapatkan. Beberapa proses validasi model terbaik meliputi validasi kecamatan, validasi pembagian data, dan validasi dengan menggunakan metode lain.

4.5.1. Validasi Kecamatan

Model terbaik pada setiap dataran akan dilakukan validasi kecamatan untuk melihat nilai akurasi model terbaik. Kecamatan yang digunakan untuk proses validasi model terbaik yaitu semua kecamatan selain kecamatan pembentuk model pada setiap dataran. Hal ini bertujuan untuk melihat model terbaik merupakan model yang *universal* atau tidak jika digunakan untuk peramalan kecamatan lain pada dataran yang sama.

4.5.2. Validasi Pembagian Data

Proses validasi pembagian data untuk model terbaik yang dihasilkan pada 3 dataran menggunakan percobaan skenario pembagian data. Skenario ini digunakan sesuai pembagian data pada pembentukan model. Tujuan dari skenario pembagian data yaitu menganalisis model terbaik bersifat *universal* atau tidak untuk beberapa proporsi data. Skenario pembagian data ditunjukkan pada Tabel 4.1.

4.5.3. Validasi Metode Lain

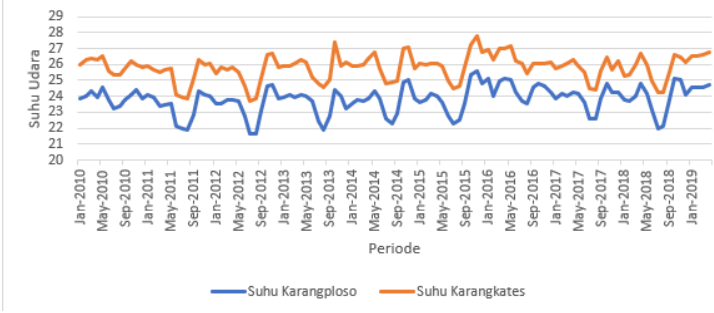
Proses validasi dengan menggunakan metode lain bertujuan untuk membandingkan model terbaik yang telah didapatkan dari metode *Random Forest Regression* dengan metode lain seperti *Decision Tree Regression, K-nearest Neighbor Regression, XG-Boost*, dan *Radial Basis Function Neural Network (RBFNN).* Metode pembanding yang digunakan memiliki cara yang berbeda-beda dalam membangun model. Metode *K-nearest Neighbor Regression* berfokus pada membangun model berdasarkan pendekatan k tetangga terdekat. Target diprediksi oleh interpolasi lokal dari target yang terkait ke k tetangga terdekat pada data *training* [22]. Sedangkan metode RBFNN merupakan metode jenis khusus dari *feed-forward neural network* yang menggunakan fungsi basis radial sebagai fungsi aktivasi. Fungsi basis radial berfungsi untuk menghubungkan antara *input, hidden layer*, dan fungsi transfer pada *hidden layer* [23].

Pada metode *XG-Boost* menggunakan prinsip peningkatan gradien untuk melengkapi model yang sudah dibangun sebelumnya. Metode ini menerapkan model formulisasi yang lebih teratur dalam mengontrol *over-fitting* [24]. Sedangkan metode *Decision Tree Regression* digunakan untuk regresi *non-parametrik.* Model yang dibangun untuk memprediksi nilai variabel target dengan cara mempelajari aturan keputusan *if-then-else*. Semakin dalam *tree* yang dibangun maka semakin kompleks aturan keputusan yang digunakan dalam membangun model [25].

4.6. Peramalan 24 Periode Mendatang

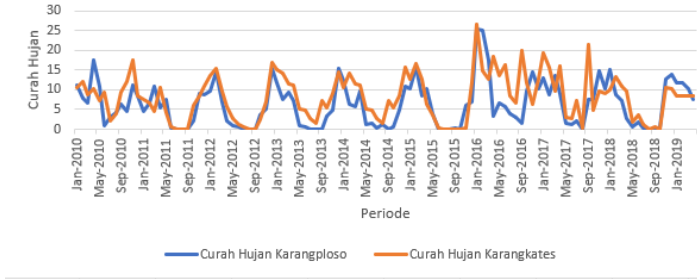
Setelah melakukan proses pengujian model dan validasi model maka didapatkan model terbaik pada setiap dataran. Model terbaik yang dihasilkan dari algoritma *Random Forest Regression* dapat digunakan untuk melakukan peramalan 24 periode mendatang. Sebelum melakukan peramalan 24 periode mendatang untuk *dependent variable*, diperlukan peramalan *independent variable* sebagai data aktual.

Peramalan *independent variable* dilakukan pada data iklim dari 2 stasiun yaitu stasiun Karangkates dan Karangploso. Peramalan *independent variable* menggunakan beberapa metode yang berbeda sesuai dengan karakteristik atau pola dari setiap data. Peramalan menggunakan *tools* Minitab, R dan Ms. Excel. Pola data suhu udara dapat ditunjukkan pada Gambar 4.1



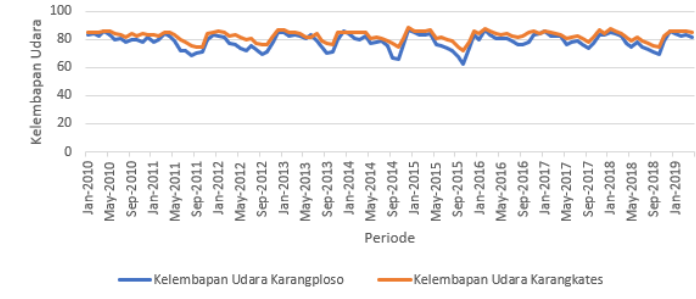
Gambar 4.1. Pola Data Suhu udara

Poal data suhu udara yang terlihat dari Gambar 4.1 merupakan pola data siklus karena memiliki fluktuasi dalam jangka yang panjang dan berulang. Metode peramalan untuk data suhu udara berupa metode Dekomposisi Multiplikatif. Untuk gambar plot 24 periode mendatang data suhu udara dapat dilihat pada Lampiran A.



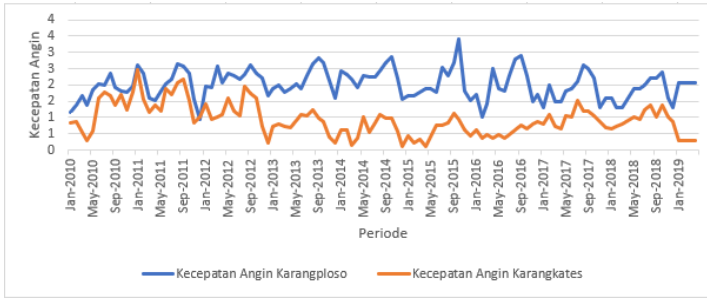
Gambar 4.2. Pola Data Curah Hujan

Pola data curah hujan yang ditunjukkan pada Gambar 4.2 berupa pola data musiman karena memiliki pola yang berulang dari waktu ke waktu. Metode yang digunakan untuk peramalan data curah hujan pada penelitian tugas akhir ini yaitu metode ARIMA. Untuk gambar plot 24 periode mendatang data curah hujan dapat dilihat pada Lampiran A.



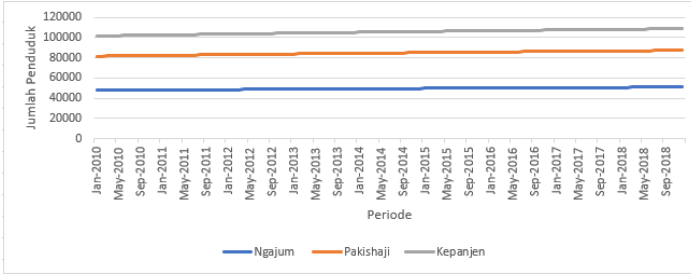
Gambar 4.3. Pola Data Kelembapan Udara

Pola data kelembapan udara yang ditunjukkan pada Gambar 4.3 berupa pola data siklus karena memiliki pola yang berulang dan fluktuasi dalam jangka panjang. Metode yang digunakan untuk peramalan data kelembapan udara yaitu metode Dekomposisi Multiplikatif. Untuk gambar plot 24 periode mendatang data kelembapan udara dapat dilihat pada Lampiran A.



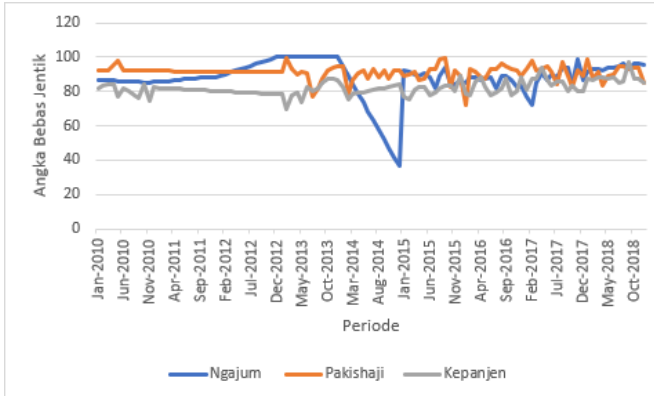
Gambar 4.4. Pola Data Kecepatan Angin

Pola data kecepatan angin yang terlihat pada Gambar 4.4 berupa pola data musiman karena memiliki pola yang berulang dari waktu ke waktu. Metode peramalan untuk data kecepatan angin yaitu metode ARIMA. Untuk gambar plot 24 periode mendatang data kecepatan angin dapat dilihat pada Lampiran A.



Gambar 4.5. Pola Data Jumlah Penduduk

Pola data jumlah penduduk pada Gambar 4.5 merupakan data jumlah penduduk 3 kecamatan di dataran tinggi. Bentuk pola data berupa *trend* karena nilai data mengalami kenaikan dalam jangka panjang. Metode yang digunakan untuk peramalan jumlah penduduk yaitu metode *Double Exponential Smoothing*. Untuk gambar plot 24 periode mendatang data jumlah penduduk dapat dilihat pada Lampiran A.



Gambar 4.6. Pola Data Angka Bebas Jentik

Pola data angka bebas jentik yang ditunjukkan pada Gambar 4.6 berupa *trend* karena memiliki kenaikan dan penuruan dalam jangka panjang. Peramalan angka bebas jentik dilakukan dengan cara menggunakan *trend forecasting* dengan metode *Exponential Smoothing* yang sederhana melalui Microsoft Excel. Untuk gambar plot 24 periode mendatang data angka bebas jentik dapat dilihat pada Lampiran A.

Selain itu, peramalan pada variabel lag dilakukan per periode. Dalam melakukan peramalan periode selanjutnya dibutuhkan variable lag sehingga variabel lag didapatkan dari variabel jumlah KDB periode terakhir. Hasil peramalan juga digunakan sebagai variabel lag untuk melakukan peramalan periode selanjutnya sesuai dengan jumlah lag yang digunakan.

Halaman ini sengaja dikosongkan

BAB V
IMPLEMENTASI

Pada bab ini akan dijelaskan mengenai penerapan metode *Random Forest Regression* dalam melakukan peramalan jumlah kasus demam berdarah di Kabupaten Malang. Urutan proses berdasarkan pada Bab Metodologi dan Bab Perancangan dengan menggunakan *tools Python.*

5.1. Persiapan Implementasi

Penelitian tugas akhir ini menggunakan perangkat keras dan lunak untuk membantu proses pengerjaannya. Spesifikasi perangkat keras dan lunak yang digunakan dapat dilihat pada Tabel 5.1.

Tabel 5.1. Spesifikasi Perangkat Keras dan Lunak

Jenis Perangkat	Laptop
Processor	Intel Core i5-7200U
Memory	8192 MB
Sistem Operasi	Windws 10 Version 1903 Build 18362.836

Proses pembentukan model peramalan menggunakan beberapa teknologi pendukung seperti *code editor*, bahasa pemograman dan *library* seperti yang ditunjukkan pada Tabel 5.2.

5.2. Pra-Processing Data

Pra-processing data dilakukan sebelum proses pembentukan model mulai dari pengolahan data iklim, *missing value*, pembagian data, uji korelasi data, normalisasi data, dan pembagian data.

5.2.1. Pengolahan Data Iklim

Data iklim yang dibutuhkan pada penilitian tugas akhir berupa data bulanan, namun data yang didapatkan dari Badan Meteorologi, Klimatologi, dan Geofisika berupa data harian. *Pra-processing* perlu dilakukan untuk mengubah data harian menjadi bulanan dengan menggunakan rata-rata seperti yang

dijelaskan pada subbab 4.2.1. Proses pengolahan data iklim dilakukan dengan menggunakan Ms. Excel.

Tabel 5.2. Teknologi Pendukung

Teknologi	Spesifikasi
Bahasa Pemograman	Python
Code Editor	Jupyter Notebook
Library	Numpy
	Matplotlib
	Plotly
	Pandas
	MinMaxScaler
	Train_test_split
	RandomForestRegressor
	Metrics
	GridSearchCV
	TimeSeriesSplit

5.2.2. Missing Value

Proses pengecekan *missing value* dilakukan pada kecamatan di setiap dataran. Pada penelitian tugas akhir ini, salah satu pengecekan *missing value* di dataran tinggi pada kecamatan Ngajum ditunjukkan pada Kode 5.1.

```
#Check missing Value
Ngajum.isnull().sum()
```

Kode 5.1. Pengecekan Missing Value

Proses pengisian *missing value* pertama pada data angka bebas jentik dimana data yang didapatkan dari Dinas Kesehatan Kabupaten Malang banyak yang kosong. Untuk pengisian *missing value* ini menggunakan 2 cara yaitu interpolasi dan regresi seperti yang dijelaskan pada subbab 4.2.2. Pengisian *missing value* dengan cara interpolasi linear yang ditunjukkan pada Kode 5.2

```
#Interpolasi data ABJ
Ngajum = Ngajum.interpolate(method ='linear', limit_di
rection ='both')
Ponco = Ponco.interpolate(method ='linear', limit_dire
ction ='both')
Jabung = Jabung.interpolate(method ='linear', limit_di
rection ='both')
```

Kode 5.2. Interpolasi

Fungsi ‘method’ bertujuan untuk memilih jenis metode interpolasi sedangkan fungsi ‘limit direction’ bertujuan untuk memilih jenis dari cara pengisian *missing value.* Penggunaan fungsi interpolasi linear dengan ‘limit direction = both’ bertujuan untuk mengisi *missing value* dari nilai antara periode sebelum dan setelahnya. Pada Kode 5.2 merupakan interpolasi linear di dataran tinggi dengan menggunakan ‘limit direction = both’.

Pengisian *missing value* dengan cara regresi dilakukan pada kecamatan Wajak di dataran sedang. Kecamatan Wajak memiliki nilai ABJ yang banyak mengandung *missing value* pada periode awal yaitu tahun 2010 sampai pertengahan tahun 2014 sehingga lebih baik menggunakan regresi dibandingkan interpolasi. Pengisian *missing value* pada kecamatan Wajak ditunjukkan pada Kode 5.3.

```
#split missing value & no missing value data ABJ
x_wajak = Wajak[Wajak['ABJ'].notnull()].drop(columns=
['ABJ'])
y_wajak = Wajak[Wajak['ABJ'].notnull()]['ABJ']
x_predict = Wajak[Wajak['ABJ'].isnull()].drop(columns
=['ABJ'])

#implementation linear regression
reg = LinearRegression().fit(x_wajak, y_wajak)
predict = reg.predict(x_predict)
wajak.ABJ[wajak.ABJ.isnull()] = predict
wajak.info()
```

Kode 5.3. Regresi

Sebelum melakukan pengisian *missing value* maka diperlukan pemisahan antara data yang memiliki *missing value* dan tidak

memiliki *missing value*. Pada Kode 5.3, variabel ‘x_wajak’ merupakan variabel yang tidak mengandung *missing value* dan variabel ‘x_predict’ merupakan variabel yang memiliki *missing value* pada kolom ABJ. Pada variabel x , kolom ABJ dihilangkan dengan tujuan kolom ABJ digunakan sebagai target untuk pengisian *missing value*. Variabel ‘y_wajak’ merupakan variabel yang hanya terdiri dari kolom ABJ dan tidak memiliki *missing value*. Pengisian *missing value* menggunakan metode linear regression. Proses pembentukan model menggunakan variabel ‘x_wajak’ dan ‘y_wajak’ sebagai *input.* Sedangkan variabel ‘x_predict’ sebagai *output.*

Pengisian *missing value* kedua pada data iklim yang memiliki beberapa nilai *missing value*. Pengisian *missing value* dilakukan dengan cara merata-rata nilai pada bulan yang sama di setiap periode seperti yang dijelaskan pada subbab 4.2.2.

5.2.3. Uji Korelasi Data

Dalam melakukan pengujian korelasi dilakukan 2 tahap uji korelasi yaitu uji korelasi terhadap pemilihan kecamatan pembentuk model dan pemilihan *independent variable*.

a. Pemilihan kecamatan pembentuk model

Proses pemilihan kecamatan dalam proses pembentukan model dilakukan pada setiap dataran. Kode program untuk melakukan uji korelasi pemilihan kecamatan terdapat pada Kode 5.4.

```
#Uji Korelasi Pemilihan Kecamatan
dataran_tinggi = d_tinggi.corr(method = 'pearson')
dataran_tinggi
```

Kode 5.4. Uji Korelasi Pemilihan Kecamatan Pembentuk Model

Fungsi ‘corr’ berguna untuk melakukan analisa korelasi secara otomatis dengan bantuan *library pandas*. Isi dari fungsi tersebut menggunakan metode jenis korelasi *pearson* yang memiliki rentang nilai -1 sampai +1.

Setelah mendapatkan hasil uji korelasi jumlah KDB antar kecamatan dalam satu dataran, pemilihan kecamatan dilakukan

berdasarkan nilai rata-rata hasil uji korelasi yang memiliki nilai tertinggi.

b. Pemilihan independent variable

Pengujian korelasi terhadap pemilihan *independent variable* dilakukan berdasarkan pada pemilihan kecamatan pembentuk model. Setelah kecamatan pembentuk model terpilih maka dilakukan pemilihan *independent variable* untuk digunakan sebagai *input*. Kode program untuk melakukan uji korelasi terhadap pemilihan variabel bebas terdapat pada Kode 5.5.

```
#Uji Korelasi variabel
corr_Ngajum = Ngajum.corr(method = 'pearson')
corr_Ngajum
```

Kode 5.5. Uji Korelasi Pemilihan Independent Variable

Proses uji korelasi pada pemilihan *independent variable* juga menggunakan fungsi ‘cor’ dengan rentang nilai -1 sampai +1. Pemilihan *independent variable* berdasarkan pada skenario pemilihan *independent variable*. Jika skenario melibatkan 3 *independent variable* maka 3 hasil uji korelasi tertinggi yang akan dipilih sebagai input dalam melakukan pembentukan model.

5.2.4. Pembagian Variabel Input dan Output

Pembagian variabel ini dilakukan sebelum proses normalisasi untuk mempermudah saat proses denormalisasi. Apabila variabel *input* dan *ouput* tidak dipisah maka akan terjadi *error* saat proses denormalisasi. Salah satu contoh pembagian variabel *input* dan *ouput* pada kecamatan Ngajum di dataran tinggi ditunjukkan pada Kode 5.6.

```
#split independent & dependent variable
x_ngajum = Ngajum[['Suhu udara', 'CH', 'JP']]
y_ngajum = Ngajum[['KDB']]
x_ngajum, y_ngajum
```

Kode 5.6. Pembagian Variabel Input dan Output

Variabel ‘x_ngajum’ berisi *independent variable* yang digunakan sebagai input. Baris pertama pada Kode 5.6 berfungsi

untuk mengambil kolom suhu udara, curah hujan, dan jumlah penduduk pada dataframe Ngajum. Variabel 'y_ngajum' berperan sebagai variabel output dimana hanya berisi kolom jumlah kasus demam berdarah.

5.2.5. Normalisasi Data

Proses normalisasi data yang digunakan pada penelitian tugas akhir ini adalah normalisasi min max dengan menggunakan fungsi 'MinMaxScaler' seperti pada Kode 5.7.

```
# normalisasi MinMaxScaler
# independent variable
scaler = MinMaxScaler(feature_range=(0,1))
xnorm_ngajum  = scaler.fit_transform(x_ngajumm)

# dependent variable
d_scaler = MinMaxScaler(feature_range=(0,1))
y_ngajum = d_scaler.fit_transform(y_ngajum)

#Dataframe train-test normalization
xnorm_ngajum = pd.DataFrame(xnorm_ngajum)
xnorm_ngajum.columns = ['Suhu udara', 'CH', 'JP']
y_ngajum = pd.DataFrame(y_ngajum)
y_ngajum.columns = ['KDB']
xnorm_ngajum, y_ngajum
```

Kode 5.7. Normalisasi Data Min Max Skenario Top 3 IV

Variabel 'scaler' dan 'd_scaler' merupakan variabel normalisasi min-max. Variabel 'scaler' digunakan untuk normalisasi *indendent variable* sedangkan variabel 'd_scaler' berperan sebagai normalisasi *dependent variable*. Fungsi 'MinMaxScaler()' berfungsi untuk melakukan normalisasi min-max dengan range 0 sampai 1.

Selain itu, terdapat fungsi 'fit_transform()' yang berguna untuk melakukan perhitungan nilai rataan terhadap data sesuai dengan rumus nomasliasi min-max. Namun, pada skenario yang hanya menggunakan variabel lag KDB maka menggunakan 1 variabel normalisasi min-max. Skenario yang hanya menggunakan variabel lag KDB mempunyai 1 kolom yaitu *dependent variable* sehingga tidak perlu dibedakan saat mendefinisikan variabel normalisasi min-max seperti pada Kode 5.8.

```
# Normalization MinMax
k_scaler = MinMaxScaler(feature_range=(0,1))
norm_ngajum = k_scaler.fit_transform(Uni_ngajum)

norm_ngajum= pd.DataFrame(norm_ngajum)
norm_ngajum.columns = ['KDB']
norm_ngajum
```

Kode 5.8. Normalisasi Min Max Skenario No IV

Variabel ‘k_scaler’ merupakan variabel normalisasi min-max untuk skenario yang hanya melibatkan variabel lag KDB. Fungsi yang digunakan sama dengan Kode 5.7 yaitu fungsi ‘MinMaxScaler()’ dengan *range* 0 sampai 1.

5.2.6. Pembentukan Lag

Dalam melakukan pembentukan lag terdapat 3 langkah yang perlu dilakukan yaitu membuat fungsi lag, membentuk variabel lag, dan menyesuaikan jumlah *dependent variable* dengan variabel lag.

1. Membuat fungsi lag

Sebelum membentuk variabel lag KDB dan menentukan jumlah lag yang akan digunakan dalam percobaan, perlu untuk membuat fungsi dalam pembentukan variabel lag. Pembuatan fungsi ini bertujuan untuk menggunakan program secara berulang kali sehingga saat membutuhkan program tersebut hanya menuliskan nama fungsi yang telah dibuat. Pembuatan fungsi lag ditunjukkan pada Kode 5.9.

```
#number of lagged
n_lag =2

#lagged function
def lags (table, n_lag, separator='_'):
    values=[]
    for i in range (n lag + 1):
        values.append(table.shift(i).copy())
        values[1].columns = [c + separator + str(i)
for c in table.columns]
    return pd.concat(values, axis=1)
```

Kode 5.9. Pembuatan Fungsi Lag

Variabel n_lag merupakan variabel yang digunakan untuk menentukan jumlah lag yang akan digunakan pada percobaan. Pada Kode 5.9 menggunakan contoh jumlah lag sebanyak 2 sehingga jika diimplementasikan akan terbentuk 2 kolom variabel lag sebagai *independent variable*.

Fungsi 'lags' yang terdapat pada baris selanjutnya merupakan nama dari fungsi dalam membentuk lag. Pada saat menjalankan fungsi tersebut maka akan terbentuk tabel baru yang berisi kolom lag. Jumlah dari kolom lag sesuai dengan pengisian angka pada variabel n_lag.

2. Membentuk variabel lag KDB

Dalam membentuk variabel lag KDB hanya perlu memanggil fungsi lags dan n_lag yang telah dibentuk pada proses sebelumnya. Setelah lag berhasil dibentuk maka perlu dilakukan penghapusan baris yang tidak memiliki nilai atau biasanya berisi 'NaN'. Baris yang tidak memiliki nilai ini terbentuk karena adanya pembentukan lag.

Jika lag sebanyak 2 maka pada kolom variabel lag 1 akan ada baris yang berisi 'NaN' sebanyak 1 dan pada kolom variabel lag 2 akan ada 'NaN' sebanyak 2. Namun, setiap kolom variabel lag tersebut akan terdapat 'NaN'. Jumlah 'NaN' yang ada di setiap kolom mengikuti jumlah lag. Penghapusan 'NaN' ini bertujuan menghindari adanya *error* saat program dijalankan. Kode program pembentukan variabel lag ditunjukkan pada Kode 5.10.

```
#build lagged KDB
lag_KDB = lags(y_ngajum[['KDB']],n_lag)
lag_KDB = lag_KDB.drop(['KDB_0'], axis =1)
lag = pd.concat([xnorm_ngajum, lag_KDB], axis =1)
x ngajum= lag.dropna()
x_ngajum
```

Kode 5.10. Pembentukan Variabel Lag

Baris pertama pada Kode 5.10 merupakan variabel pembentukan lag dengan memanggil fungsi 'lags'. Lag yang akan dibentuk menggunakan data kecamatan Ngajum dengan

kolom KDB. Variabel ‘n_lag’ merupakan variabel yang mewakili jumlah lag yang akan dibentuk.

Variabel ‘x_ngajum’ merupakan variabel yang berisi gabungan antara *independent variable* yang lain (suhu udara, jumlah penduduk, curah hujan, kelembapan udara, kecepatan angin, dan angka bebas jentik) dengan variabel lag yang sudah dibentuk. Pada kondisi ini, data masih mengandung ‘NaN’ sehingga perlu dihilangkan. Variabel ‘x_ngajum’ merupakan variabel yang akan digunakan dalam melakukan percobaan karena ‘NaN’ yang berada di variabel lag sudah dihapus dengan cara melakukan drop pada baris yang memiliki nilai ‘NaN’. Dampak menghilangkan ‘NaN’ ini yaitu data akan kehilangan baris sebanyak jumlah lag yang dibentuk.

3. Menyesuaikan jumlah dependent variable dengan variable lag

Jumlah periode atau baris antara x sebagai input dan y sebagai target harus sama untuk menghindari *error* saat program dijalankan. Untuk menyesuaikam jumlah antara *independent variabel* lag dengan *dependent variable* maka perlu adanya pengambilan baris sesuai dengan jumlah lag seperti yang ditunjukkan pada Kode 5.11.

```
#dependent variable
y_ngajum = y_ngajum.loc[n_lag:]
y_ngajum
```

Kode 5.11. Pengambilan Baris Dependent Variable sesuai n_lag

Baris pertama pada Kode 5.11 berguna untuk mengambil baris *dependent variable* sesuai dengan jumlah lag yang telah ditentukan sebelumnya.

5.2.7. Pembagian Data Training dan Testing

Pembagian data pada penelitian tugas akhir ini berdasarkan pada 4 skenario seperti pada Tabel 4.1. Jumlah data pelatihan dan pengujian dipengaruhi oleh jumlah lag yang dibentuk sehingga setiap percobaan lag akan dihasilkan jumlah data yang berbeda-

beda. Kode program untuk melakukan pembagian data terdapat pada Kode 5.12.

```
#split train dan test  x and y
x_train, x_test, y_train, y_test = train_test_split(x_
ngajum,y_ngajum,test_size=0.3,shuffle=False)
x_train.shape, x_test.shape
```

Kode 5.12. Pembagian Data Training dan Data Testing

Pada saat akan melakukan perubahan terhadap skenario pembagian data di setiap percobaan, variabel yang dapat diubah yaitu variabel ‘test_size’. Pada Kode 5.12, data dibagi menjadi 70% data training dan 30% data testing karena pada variabel ‘test_size’ berisi angka 0.3.

5.3. Pembentukan Model Random Forest

Proses pembentukan model dengan menggunakan metode *Random Forest Regression* diperlukan 2 langkah inisiasi parameter dan proses pemodelan.

5.3.1. Inisiasi Parameter

Proses yang dilakukan sebelum pembentukan model yaitu inisiasi nilai dari parameter. Parameter yang digunakan pada penelitian tugas akhir ini sebanyak 3 yaitu jumlah n_estimator, jumlah min_samples_split, dan jumlah max_depth. Penentuan nilai 3 parameter seperti pada Tabel 4.2. Kode program yang digunakan untuk inisiasi parameter terdapat pada Kode 5.13.

```
#parameter RFR
parameter = {
    'max_depth': [5, 10, 15, 20],
    'min_samples_split': [4, 8, 12, 16],
    'n_estimators': [100, 200, 300, 400, 500,600,700,8
00,900,1000]
}
```

Kode 5.13. Nilai Parameter

Variabel parameter pada Kode 5.13 berperan sebagai *array* yang berisi nilai dari 3 parameter yang digunakan dalam pembentukan model.

5.3.2. Pemodelan

Proses pembentukan model menggunakan *library RandomForestRegressor* yang terdapat pada bahasa pemograman *python*. Selain itu, penelitian tugas akhir ini memanfaatkan *library GridSearchCV*. Penerapan *GridSearchCV* dapat membantu dalam melakukan pemilihan parameter terbaik yang dihasilkan pada setiap percobaan. Kode program untuk pembentukan model ditunjukkan pada Kode 5.14.

```
# implementation Grid Search

tscv = TimeSeriesSplit(n_splits=5)
rf = RandomForestRegressor(random_state = 42)
grid_search = GridSearchCV(estimator = rf, param_grid
= parameter,cv = tscv, n_jobs = -1,verbose =1)

# build model
grid_search.fit(x_train, y_train)
grid_search.best_params_
```

Kode 5.14. Pembentukan Model

Penerapan *Cross Validation* pada *GridSearchCV* memanfaatkan *library TimeSeriesSplit* bukan *k-fold cross-validation*. Data yang tergolong *time series* lebih cocok menggunakan *time series split cross-validation*. Proses split data *training* dan data *testing* pada *times series split cross-validation* tidak acak sedangkan pada *k-fold cross-validation* dilakukan secara acak. Data testing pada *k-fold cross-validation* dapat terletak diantara data *training* ataupun setelah data *training* sehingga memungkinkan ada 3 *split*. Sedangkan pada *times series split cross-validation*, data testing diambil setelah data training sehingga proses *split* dilakukan hanya 2 kali [26].

Variabel ‘rf’ merupakan variabel yang berisi *library RandomForestRegression* dan menggunakan ‘random_state = 42’. Baris ketiga pada Kode 5.14 merupakan variabel ‘grid_search’ yang berisi *library*, parameter dan penerapan *time series split cross-validation*. Variabel ‘grid_search’ ini yang akan digunakan dalam pembentukan model. Setelah ditemukan

parameter terbaik dari pembentukan model pada data *training*, proses selanjutnya dilakukan pengujian dengan menggunakan data *testing* yang ditunjukkan pada Kode 5.15.

```
best_grid = grid_search.best_estimator_
#Train_Pred
ytrain_pred = best_grid.predict(x_train).reshape(-1,1)
ytrain_pred_denorm = d_scaler.inverse_transform(ytrain
_pred)

#Test_Pred
ytest_pred = best_grid.predict(x_test).reshape(-1,1)
ytest_pred_denorm = d_scaler.inverse_transform(ytest_p
red)

y_test_denorm = d_scaler.inverse_transform(y_test)
y_train_denorm = d_scaler.inverse_transform(y_train)
```

Kode 5.15. Penerapan Model Terbaik Model pada Data Testing

Baris pertama pada Kode 5.15 merupakan variabel ‘best_grid’ yang digunakan untuk memanggil fungsi model terbaik yang didapatkan saat pembentukan model. Variabel ‘ytest_pred’ merupakan variabel yang akan menghasilkan data hasil prediksi dengan menggunakan ‘x_test’ sebagai data testing. Variabel ‘ytest_pred_denorm’ berisi data hasil prediksi yang telah didenormalisasi sedangkan variabel ‘y_test_denorm’ berisi data aktual yang telah didenormalisasi.

5.4. Pengujian Model Random Forest

Pengujian model yang dihasilkan dari proses pembentukan model berdasarkan pada nilai RMSE. Untuk proses validasi model terhadap kecamatan lain dan pembagian data digunakan nilai SMAPE. Perhitungan nili RMSE memanfaatkan *library metrics* sehingga tidak perlu mendefinisikan rumus RMSE sedangkan untuk perhitungan nilai SMAPE diperlukan pendefinisian rumus. Kode program untuk rumus SMAPE ditunjukkan Kode 5.16.

```
#Rumus SMAPE
def smape (A, F):
    return 100/len(A) * np.sum(np.abs(F-
A)/ (np.abs(A) + np.abs(F)))
```

Kode 5.16. Mendefinisikan Rumus SMAPE

Untuk mencetak nilai *error* dari RMSE dan SMAPE antara data aktual dengan data hasil prediksi ditunjukkan pada Kode 5.17.

```
#Print Error
print('RMSE-
Training:', np.sqrt(metrics.mean_squared_error
                (y_train_denorm, ytrain_pred_denorm)))
print('RMSE-
Testing:', np.sqrt(metrics.mean_squared_error
                 (y_test_denorm, ytest_pred_denorm)))
print("SMAPE-
Testing:" , smape(y_test_denorm, ytest_pred_denorm))
```

Kode 5.17. Perhitungan Nilai RMSE dan SMAPE

5.5. Validasi Model

Validasi model dilakukan terhadap model terbaik yang telah dihasilkan pada proses pembentukan model. Pada proses ini dilakukan 2 validasi yaitu validasi kecamatan lain dan validasi pembagian data.

5.5.1. Validasi Kecamatan Lain

Proses validasi pertama yaitu terhadap kecamatan lain. Kecamatan yang digunakan untuk proses validasi model terbaik yaitu semua kecamatan selain kecamatan pembentuk model pada setiap dataran. Kode program validasi model terbaik terhadap kecamatan lain ditunjukkan pada Kode 5.18.

```
#melakukan proses normalisasi MinMax
# normalization on independent
ponco_scaler = MinMaxScaler(feature_range=(0,1))
x_ponco  = ponco_scaler.fit_transform(x_ponco)

# normalization on dependent
dp scaler = MinMaxScaler(feature range=(0,1))
y_ponco = dp_scaler.fit_transform(y_ponco)

#Dataframe train-test normalisasi
x_ponco = pd.DataFrame(x_ponco)
x_ponco.columns = ['Suhu udara', 'CH', 'JP']
y_ponco = pd.DataFrame(y_ponco)
y_ponco.columns = ['KDB']
x_ponco, y_ponco
```

Kode 5.18.Proses Normalisasi pada Kecamatan Lain

Dalam melakukan proses validasi model terbaik terhadap kecamata lain, kode program yang digunakan sama dengan kode dalam pembentukan model. Namun, variabel ‘scaler’ yang digunakan pada setiap kecamatan validasi harus berbeda antara satu dengan yang lain. Pada Kode 5.18 menggunakan data kecamatan Poncokusumo dan variabel ‘dp_scaler’ sebagai variabel ‘scaler’ untuk proses normalisasi pada kecamatan validasi. Pada proses validasi kecamatan tidak melakukan pembagian data *training* dan data *testing.*

5.5.2. Validasi Pembagian Data

Validasi pembagian data dilakukan terhadap model terbaik yang telah dihasilkan dari semua skenario di setiap dataran. Proses ini menggunakan pembagian data berdasarkan pada Tabel 4.1.

Jika pembentukan model menggunakan 70% data *training* dan 30% data *testing*, maka terdapat 3 bentuk validasi pembagian data yaitu 80% data *training* dan 20% data *testing*, 60% data *training* dan 40% data *testing* serta 50% data *training* dan 50% data *testing*. Validasi dilakukan dengan menggunakan model terbaik dari semua skenario *independent variable* di pembentukan model dengan pembagian data 70% data *training* dan 30% data *testing*. Kode program untuk melakukan validasi pembagian data ditunjukkan pada Kode 5.19.

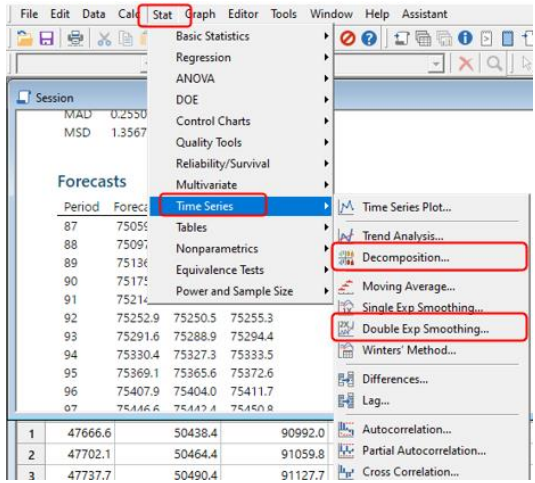
```
#split train dan test  x and y
x train, x test, y train, y test = train test split(x
ngajum,y_ngajum,test_size=0.5,shuffle=False)
x_train.shape, x_test.shape
```

Kode 5.19. Validasi Pembagian Data

Dalam melakukan validasi data, kode program yang digunakan sama dengan proses pembagian data untuk pemodelan. Pada Kode 5.19, variabel ‘test_size’ bernilai 0.5 dimana skenario validasi pembagian data yang dilakukan yaitu 50% data *training* dan 50% data *testing*. Untuk menerapkan model terbaik pada kecamatan lain dapat menggunakan kode program seperti pada Kode 5.15.

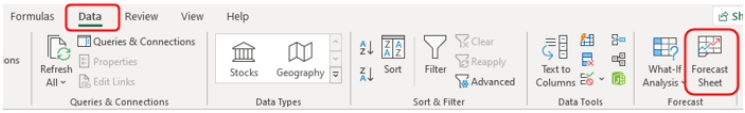
5.6. Peramalan 24 Periode Mendatang

Proses peramalan 24 periode mendatang digunakan untuk meramalkan jumlah KDBsebagai *dependent variable* atau *output*. Peramalan dilakukan pada kecamatan pembentuk model dan kecamatan lain di setiap dataran berdasarkan model terbaik. Sebelum melakukan peramalan 24 periode mendatang untuk *dependent variable*, diperlukan peramalan *independent variable* sebagai data aktual. Peramalan data suhu udara, jumlah penduduk dan kelembapan udara menggunakan tools Minitab dengan cara pilih menu Stat → Time Series → Decomposition/ Double Exp. Smoothing. Metode Decomposition untuk meramalkan data suhu udara dan kelembapan udara. Sedangkan metode Double Exp. Smoothing untuk meramalkan data jumlah penduduk. Proses peramalan dengan menggunakan tools Minitab ditunjukkan pada Gambar 5.1.



Gambar 5.1. Peramalan menggunakan Minitab

Peramalan data Angka Bebas Jentik (ABJ) menggunakan tools Ms. Excel dengan cara pilih menu Data → Forecast Sheet seperti yang ditunjukkan pada Gambar 5.2 . Sedangkan untuk meramalkan data curah hujan dan kecepatan angin menggunakan tools R yang ditunjukkan pada Kode 5.20 .



Gambar 5.2. Peramalan menggunakan Ms. Excel

```
CH_KK <- read_excel("F:/TA FIRINHAN/CH_Karangkates.xls
x")
CH_KK.ts <- ts(CH_KK, start = c(2010,1), frequency = 1
2)
plot(CH_KK.ts)

components.ts = decompose(CH_KK.ts)
plot(components.ts)
model = auto.arima(CH_KK.ts, trace=TRUE)

predict(model,n.ahead = 24)
peramalan = forecast(model,24)
plot(forecast(model,24))
forecast_CH <- data.frame(peramalan)
```

Kode 5.20. Peramalan data curah hujan menggunakan R

Peramalan 24 periode mendatang untuk variabel lag dilakukan dengan cara menggunakan nilai KDB pada periode terakhir yaitu bulan Desember 2018. Nilai tersebut akan digunakan untuk data *input* yang berperan sebagai variabel lag. Nilai KDB sementara diisi dengan nilai 0 sehingga dalam pemograman tidak terbaca 'NaN'. Sedangkan saat meramalkan periode 2020 maka variabel lag yang digunakan yaitu hasil peramalan jumlah KDB pada periode 2019.

Sebelum melakukan peramalan 24 periode mendatang, pembuatan fungsi peramalan perlu dilakukan. Tujuan pembuatan fungsi dalam bentuk kode program untuk mempermudah proses peramalan 24 periode mendatang sehingga tidak perlu dilakukan satu per satu setiap periode secara manual. Kode progam dalam meramalkan 24 periode antara melibatkan hanya variabel lag dan semua *independent variable* termasuk variabel lag dibuat berbeda. Proses peramalan jumlah KDB untuk 24 periode mendatang hanya dengan melibatkan variabel lag ditunjukkan pada Kode 5.21.

```
#variable lag without IV
#forecast next period
def forecastNextPeriod(n_period_tobeforecasted, data, d
ata_ori):
    current_nextperiod = 1
    # initiate scaler object
    x_scaler = MinMaxScaler(feature_range=(0,1))

while current_nextperiod <= n_period_tobeforecasted:
Uni_feature = data[['KDB']] # with next period
Uni_feature = Uni_feature.fillna(value = 0)

# normalization 108 periode
Uni_feature_normalized  = x_scaler.fit_transform(Uni_fe
ature[:(len(data ori)+current nextperiod-1)])
Uni_feature_normalized = pd.DataFrame(Uni_feature_norma
lized)
Uni_feature_normalized.columns = ['KDB']

# normalization next period
Uni feature normalized = np.append(Uni feature normaliz
ed.values, 0)
Uni_feature_normalized = pd.DataFrame(Uni_feature_norma
lized)
Uni_feature_normalized.columns = ['KDB']

# lagged KDB
lag_KDB = lags(Uni_feature_normalized.tail(4), n_lag)
lag_KDB = lag_KDB.drop(['KDB_0'], axis =1)
x = lag_KDB.dropna()

# forecasted next period
current_nextperiod_pred = best_grid.predict(x).reshape(
-1,1)
current_nextperiod_pred_denorm = x_scaler.inverse_trans
form(current_nextperiod_pred)

# replace 0 value with forecasted
data.loc[((len(data_ori)-
1)+current_nextperiod), 'KDB'] = current_nextperiod_pre
d_denorm[-1]
current_nextperiod+=1
return data
```

Kode 5.21. Peramalan Jumlah KDBdengan Melibatkan hanya Variabel Lag

Kode 5.21 merupakan *syntax* untuk melakukan peramalan jumlah kasus demam berdarah yang hanya melibatkan variabel lag. Variabel 'forecastNextPeriod' merupakan fungsi yang akan dipanggil saat melakukan peramalan sebanyak n periode.

Fungsi ini berisi jumlah n periode, data gabungan antara aktual dengan dataset yang ada penambahan baris untuk periode selanjutnya, dan data aktual yang berisi jumlah KDB sebanyak 108 periode. Proses normalisasi jumlah KDB untuk periode selanjutnya mengikuti *object scaler* normalisasi variabel x yang digunakan saat pembentukan model sehingga *range* nilainya sama.

Untuk peramalan jumlah KDB 24 periode mendatang yang melibatkan *independent variable* dan variabel lag ditunjukkan pada Kode 5.22.

Kode 5.22 merupakan *syntax* untuk mendefinisikan fungsi 'forecastNextPeriod' dan proses normalisasi . Hal yang membedakan dengan Kode 5.21 yaitu proses normalisasi pada Kode 5.22 melibatkan y dimana *object scaler* yang digunakan mengikuti *range* dari normalisasi jumlah KDB pada saat pembentukan model. *Object scaler* untuk variabel x menggunakan *range* dari normalisasi *independent variable* saat pembentukan model. Fungsi peramalan yang hanya melibatkan variabel lag memiliki kode program lebih pendek karena hanya ada 1 *object scaler* saja.

Kode 5.23 merupakan kode program yang dijalankan untuk melakukan peramalan 24 periode mendatang. Salah satu contoh yang ditunjukkan pada Kode 5.23 yaitu meramalkan jumlah KDB untuk 24 periode mendatang pada kecamatan Ngajum.

```
#with All independent variable & variable lag
#forecast next period
def forecastNextPeriod(n_period_tobeforecasted, data, data_or
i):
current_nextperiod = 1
#split independent & dependent variable
independent_feature = data[['Suhu udara', 'CH', 'JP']]
# normalization on independent variable
x_scaler = MinMaxScaler(feature_range=(0,1))
# normalization on dependent variable
y_scaler = MinMaxScaler(feature_range=(0,1))

while current_nextperiod <= n_period_tobeforecasted:
# normalization on independent
# normalization before forecasted (108 period)
xnorm_hingga_current_nextperiod  = x_scaler.fit_transform(ind
ependent_feature[:(len(data_ori)+current_nextperiod)])
# xnorm_hingga_current_nextperiod  = x_scaler.fit_transform(i
ndependent_feature[:(len(data_ori)+1)])
xnorm_hingga_current_nextperiod = pd.DataFrame(xnorm_hingga_c
urrent nextperiod)
xnorm_hingga_current_nextperiod.columns = ['Suhu
udara', 'CH', 'JP']
# normalization on forecasted independent (109-132 period)
xnorm_stelahperiode108 = x_scaler.transform(independent_featu
re[len(data_ori):])
xnorm stelahperiode108 = pd.DataFrame(xnorm stelahperiode108)
xnorm_stelahperiode108.columns = ['Suhu udara', 'CH', 'JP']
# join independent 108 period + next period
x_withnextperiode = pd.concat((xnorm_hingga_current_nextperio
d, xnorm_stelahperiode108), axis = 0)
x_withnextperiode = x_withnextperiode.reset_index(drop= True)

# normalization on independent
# fill nan value in 109-132 period to 0
y_withnextperiode = data[['KDB']].fillna(value = 0)
# normalisasi dependen periode 1-108
y_withoutnextperiode_norm = y_scaler.fit_transform(y_withnext
periode[:(len(data_ori)+current_nextperiod)])
y_withoutnextperiode_norm = pd.DataFrame(y_withoutnextperiode
_norm)
y_withoutnextperiode_norm.columns = ['KDB']
# combine normalized dependent with next periode
y_data = pd.concat((y_withoutnextperiode_norm, y_withnextperi
ode[(len(data_ori)+current_nextperiod):]),axis = 0)
```

(a)

```
# lagged KDB
lag_KDB = lags(y_data[0:(len(data_ori)+current_nextper
iod)].tail(24), n_lag)
lag_KDB = lag_KDB.drop(['KDB_0'], axis =1)
data_withlag = pd.concat([xnorm_hingga_current_nextper
iod, lag_KDB], axis =1)
x_data = data_withlag.dropna()

# forecasted next period
current_nextperiod_pred = best_grid.predict(x_data).re
shape(-1,1)
current_nextperiod_pred_denorm = y_scaler.inverse_tran
sform(current_nextperiod_pred)
# replace 0 value with forecasted
data.loc[((len(data ori)-
1)+current_nextperiod), 'KDB'] = current_nextperiod_pr
ed_denorm[11]
current_nextperiod+=1
return data
```

(b)

Kode 5.22. Melakukan Normalisasi (a) dan Mendefinisikan Fungsi Peramalan Periode Mendatang (b)

```
#forecasted 24 period
n_period_tobeforecasted = 24
future_KDB = forecastNextPeriod(n_period_tobeforecaste
d, D_Ngajum, Ngajum)
future_KDB[107:]
```

Kode 5.23. Peramalan Jumlah KDB24 Periode Mendatang

Halaman ini sengaja dikosongkan

BAB VI
HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan mengenai hasil dari uji coba dan pembahasan serta dilakukan proses analisis terhadap hasil yang diperoleh dari proses pelatihan dan pengujian model dengan metode *Random Forest Regression*.

6.1 Hasil Pra-Processing Data

Hasil dari implementasi pra-proses data berkaitan dengan perubahan data iklim, data angka bebas jentik, uji korelasi dalam pemilihan kecamatan dan *independent variable* serta pembentukan lag jumlah KDB.

6.1.1 Hasil Pengolahan Data Harian Iklim

Data harian iklim yang dilakukan pengolahan yaitu suhu udara, curah hujan, kecepatan angin, dan kelembapan udara. Salah satu contoh dalam melakukan proses pengolahan data iklim di dataran rendah pada bulan Januari 2010 yang ditunjukkan pada Tabel 6.1.

Pada Tabel 6.1, terdapat 5 kolom yang berisi tanggal, variabel Tavg (rata-rata suhu udara), variabel RH_avg (rata-rata kelembapan udara), variabel Ff_avg (rata-rata kecepatan angin), dan variabel RR (rata-rata curah hujan). Untuk hasil pengolahan data harian iklim dengan cara menghitung rata-rata untuk bulan Januari 2010 ditunjukkan pada Tabel 6.2.

6.1.2 Hasil Analisa Missing Value

Proses yang dilakukan pada analisis *missing value* yaitu pengecekan dan pengisian *missing value*. Hasil pengecekan terhadap *missing value* ditunjukkan pada Tabel 6.3.

Tabel 6.1. Data Harian Iklim Bulan Januari 2010 pada Dataran Rendah

Tanggal	Tavg	RH_avg	Ff_avg	RR
01-01-2010	26.2	86	1	1
02-01-2010	26.9	85	0	10
03-01-2010	27.3	84	1	6
04-01-2010	27.2	84	1	10
05-01-2010	26.3	83	0	1
06-01-2010	26.8	83	0	0
07-01-2010	26.4	86	0	0
08-01-2010	25.7	88	0	15
09-01-2010	24.2	92	0	12
10-01-2010	25.6	88	0	7
11-01-2010	25.9	81	2	2
12-01-2010	26.2	85	2	20
13-01-2010	25.7	83	2	44
14-01-2010	25.7	83	1	1
15-01-2010	26.3	76	1	6
16-01-2010	26	94	1	2
17-01-2010	26	80	1	8
18-01-2010	27.3	75	1	0
19-01-2010	26.6	84	1	5
20-01-2010	25.7	86	1	2
21-01-2010	24.8	89	1	52
22-01-2010	25.3	88	0	4
23-01-2010	26.9	82	0	0
24-01-2010	26.2	85	0	0
25-01-2010	25.6	84	2	19
26-01-2010	25.4	84	1	90
27-01-2010	25.3	86	2	1
28-01-2010	25.1	88	1	2
29-01-2010	26.2	85	1	5
30-01-2010	25.8	86	1	0
31-01-2010	26.1	84	1	1

Tabel 6.2. Hasil Pengolahan Data Iklim Bulan Januari 2010

Temperature	Humidity	Wind Speed	Rainfall
26.02	84.74	0.84	10.52

Tabel 6.3. Hasil Pengecekan Missing Value di Kecamatan Ngajum

Periode	0
Suhu udara	6
KU	4
CH	7
KA	0
JP	0
ABJ	43
KDB	0

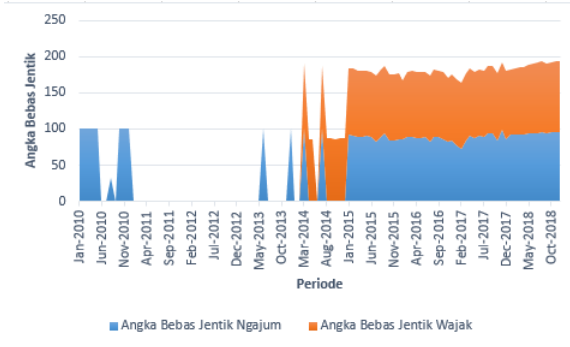
Pada Tabel 6.3 dapat terlihat bahwa data Angka Bebas Jentik (ABJ) terdapat 43 *missing value* sehingga perlu adanya pra-proses data. Pengisian *missing value* dilakukan dengan cara interpolasi dan regresi. Interpolasi diterapkan pada kecamatan yang tidak banyak memiliki *missing value* seperti kecamatan Ngajum di dataran tinggi yang ditunjukkan pada Tabel 6.4.

Tabel 6.4. Data Angka Bebas Jentik di Kecamatan Ngajum Tahun 2010

Periode	ABJ	Periode	ABJ
Jan-2010	100	Jul-2010	
Feb-2010	100	Agt-2010	33.33333
Mar-2010	100	Sep-2010	0
Apr-2010	100	Okt-2010	100
Mei-2010	100	Nop-2010	100
Jun-2010		Des-2010	100

Kecamatan Ngajum memiliki data ABJ kosong di pertengahan periode sehingga masih bisa menggunakan interpolasi. Sedangkan pada kasus kecamatan Wajak di dataran sedang, pengisian *missing value* dilakukan dengan cara regresi linear

karena data ABJ dari periode Januari 2010 sampai Februari 2014 kosong. Perbandingan grafik *missing value* antara kecamatan Ngajum dan Wajak ditunjukkan pada Gambar 6.1.



Gambar 6.1. Perbandingan missing value kecamatan Ngajum dan Wajak

Pada Gambar 6.1 terlihat bahwa dari tahun 2010 sampai 2014 data angka bebas jentik pada kecamatan Wajak kosong sehingga jika dilakukan pengisian *missing value* dengan interpolasi maka akan memiliki nilai yang sama atau tidak bervariasi. Selain data ABJ, pengisian *missing value* dilakukan pada data iklim dengan cara melakukan rata-rata pada bulan yang sama di setiap periode. Hasil pengecekan setelah dilakukan pengisian *missing value* ditunjukkan pada Tabel 6.5.

Tabel 6.5. Hasil Pengisian Missing Value Data ABJ di Kecamatan Ngajum Tahun 2010

Periode	0
Suhu udara	0
KU	0
CH	0
KA	0
JP	0
ABJ	0
KDB	0

6.1.3 Hasil Uji Korelasi Pemilihan Kecamatan Pembentuk Model

Pemilihan kecamatan pembentuk model dilakukan pada setiap dataran dengan berdasarkan rata-rata nilai uji korelasi. Salah satu contoh hasil uji korelasi di kecamatan Ngajum ditunjukkan pada Tabel 6.6.

Tabel 6.6. Hasil Korelasi antarkecamatan di Dataran Tinggi

Kecamatan	Poncokusumo	Jabung	Ngajum
Poncokusumo	1	0.2834	0.6879
Jabung	0.2834	1	0.4794
Ngajum	0.6879	0.4794	1
avg	0.4856	0.3814	0.5836

Setelah didapatkan hasil uji korelasi antarkecamatan di dataran tinggi maka dilakukan rata-rata setiap hasil korelasi dengan cara manual di Ms. Excel. Hasil dari rata-rata uji korelasi di dataran tinggi menunjukkan bahwa kecamatan Ngajum memiliki korelasi yang kuat terhadap kecamatan lainnya dengan nilai rata-rata uji korelasi sebesar 0.583622. Pada dataran tinggi, kecamatan yang terpilih sebagai pembentuk model yaitu kecamatan Ngajum.

Proses uji korelasi antarkecamatan juga dilakukan pada dataran sedang dan dataran rendah. Pada dataran sedang kecamatan yang terpilih sebagai pembentuk model yaitu kecamatan Pakishaji dengan nilai rata-rata uji korelasi sebesar 0.371269. Pada dataran rendah, kecamatan Kepanjen terpilih sebagai kecamatan pembentuk model dengan nilai rata-rata uji korelasi sebesar 0.632633. Untuk nilai hasil uji korelasi antarkecamatan di dataran rendah dan sedang dapat dilihat di Lampiran B.

6.1.4 Hasil Uji Korelasi Pemilihan Independent Variable

Pada sub bab ini dijelaskan hasil uji korelasi *independent variable* terhadap *dependent variable* pada kecamatan pembentuk model di setiap dataran. Proses ini dilakukan di setiap kecamatan pembentuk model. Hasil uji korelasi digunakan sebagai dasar pemilihan *independent variable* sesuai

dengan skenario *independent variable* untuk pembentukan model seperti yang ditunjukkan pada Tabel 4.3 sub bab 4.3.2.

a. Dataran Tinggi

Pada dataran tinggi, uji korelasi pemilihan *independent variable* di kecamatan Ngajum sebagai kecamatan pembentuk model. Hasil uji korelasi *independent variable* di kecamatan Ngajum dapat dilihat pada Tabel 6.7.

Tabel 6.7. Hasil Uji Korelasi Pemilihan Independent Variable di Kecamatan Ngajum

	Suhu udara	KU	CH	KA	ABJ	JP	KDB
Suhu udara	1						
KU	0.2077	1					
CH	0.4468	0.6600	1				
KA	-0.4534	-0.4997	-0.4708	1			
ABJ	-0.0853	-0.0777	-0.0549	0.0022	1		
JP	0.1078	-0.0775	0.0781	-0.3735	-0.1904	1	
KDB	0.1060	0.0870	0.2953	-0.0907	-0.0873	0.2789	1

Pada tabel 6.7 terlihat bahwa 3 *independent variable* yang memiliki nilai korelasi paling kuat terhadap jumlah KDB yaitu variabel CH (curah hujan), variabel JP (jumlah penduduk), dan variabel Suhu udara. Pembentukan model dengan menggunakan skenario Top 3 IV dapat menjadi salah satu contoh penggunaan 3 *independent variable* yang memiliki 3 nilai korelasi paling kuat tersebut. Untuk skenario Top 2 IV melibatkan variabel CH dan JP karena memiliki nilai korelasi 2 teratas sedangkan pada skenario Top 1 IV hanya melibatkan variabel CH karena memiliki nilai korelasi paling kuat diantara yang lain.

b. Dataran Sedang

Pada dataran sedang, uji korelasi pemilihan *independent variable* menggunakan kecamatan Pakishaji sebagai kecamatan pembentuk model. Hasil uji korelasi *independent variable* di kecamatan Pakishaji dapat dilihat pada Tabel 6.8.

Tabel 6.8. Hasil Uji Korelasi Independent Variable di Kecamatan Pakishaji

	Suhu udara	KU	CH	KA	ABJ	JP	KDB
Suhu udara	1						
KU	0.2077	1					
CH	0.4468	0.6600	1				
KA	-0.4534	-0.4997	-0.4708	1			
ABJ	-0.1304	0.0097	-0.1413	0.0655	1		
JP	0.1078	-0.0775	0.0781	-0.3734	-0.1044	1	
KDB	0.0376	0.1915	0.2047	-0.0446	-0.2850	-0.1328	1

Dari Tabel 6.8 dapat diketahui bahwa *independent variable* yang memiliki 3 nilai tertinggi yaitu variabel ABJ (angka bebas jentik), variabel CH (curah hujan), dan variabel KU (kelembapan udara).

Dalam pembentukan model dengan menggunakan skenario Top 3 IV dapat melibatkan variabel ABJ, CH, dan KU. Untuk skenario Top 2 IV melibatkan variabel ABJ dan CH karena memiliki nilai korelasi 2 teratas sedangkan pada skenario Top 1 IV hanya melibatkan variabel ABJ karena memiliki nilai korelasi tertinggi.

c. Dataran Rendah

Pada dataran rendah, uji korelasi pemilihan *independent variable* di kecamatan Kepanjen sebagai kecamatan pembentuk model. Hasil uji korelasi *independent variable* di kecamatan Kepanjen dapat dilihat pada Tabel 6.9.

Pada tabel 6.9 terlihat bahwa 3 *independent variable* yang memiliki nilai korelasi paling kuat terhadap jumlah KDB yaitu variabel variabel JP (jumlah penduduk), variabel KU (kelembapan udara) dan variabel CH (curah hujan).

Tabel 6.9. Hasil Uji Korelasi Pemilihan Independent Variable di Kecamatan Kepanjen

	Suhu udara	KU	CH	KA	ABJ	JP	KDB
Suhu udara	1						
KU	0.2077	1					
CH	0.4468	0.6600	1				
KA	-0.4534	-0.4997	-0.4708	1			
ABJ	0.0730	-0.1527	-0.0588	-0.0325	1		
JP	0.1078	-0.0775	0.0781	-0.3735	0.4598	1	
KDB	0.0509	0.2357	0.2094	0.1244	-0.1078	-0.2761	1

Pembentukan model dengan menggunakan skenario Top 3 IV menggunakan 3 *independent variable* yang memiliki 3 nilai korelasi paling kuat tersebut. Untuk skenario Top 2 IV melibatkan variabel JP dan KU karena memiliki nilai korelasi 2 teratas sedangkan pada skenario Top 1 IV hanya melibatkan variabel JP karena memiliki nilai korelasi paling kuat diantara yang lain.

6.1.5 Hasil Pembagian Variabel Input dan Output

Proses pembagian variabel *input* dan *output* digunakan untuk memisahkan antara *independent variable* sebagai input dan *dependent variable* sebagai output. Salah satu hasil pemisahan variabel input pada kecamatan Ngajum di dataran tinggi yang ditunjukkan pada Tabel 6.10

Tabel 6.10 merupakan salah satu contoh dari hasil pemisahan *independent variabel* namun masih belum melibatkan variabel lag. Periode yang ditunjukkan pada Tabel 6.10 hanya 12 periode saja yaitu tahun 2010 sedangkan dalam pembentukan model terdapat 108 periode yaitu dari tahun 2010 sampai tahun 2018. Untuk hasil pembagian dependent variable ditunjukkan pada Tabel 6.11.

Tabel 6.10. Hasil Pemisahan Independent Variable sebagai Input

Periode	Suhu	CH	JP
Jan-2010	23.85	83.06	11.32
Feb-2010	23.99	83.82	7.79
Mar-2010	24.36	82.55	6.74
Apr-2010	23.91	85.50	17.63
Mei-2010	24.59	82.81	11.00
Jun-2010	23.79	79.57	1.00
Jul-2010	23.21	80.58	2.97
Agt-2010	23.39	77.87	4.35
Sep-2010	23.80	80.17	6.27
Okt-2010	24.09	79.81	4.55
Nop-2010	24.42	78.00	11.13
Des-2010	23.83	81.87	8.42

Tabel 6.11. Hasil Pemisahan Dependent Variable sebagai Output

Periode	KDB	Periode	KDB
Jan-2010	1	Jul-2010	0
Feb-2010	3	Agt-2010	0
Mar-2010	0	Sep-2010	0
Apr-2010	1	Okt-2010	0
Mei-2010	0	Nop-2010	0
Jun-2010	0	Des-2010	0

Pada Tabel 6.11 menunjukkan jumlah KDB sebagai *dependent variable* dalam proses pembentukan model pada kecamatan Ngajum di dataran tinggi. Periode yang ditunjukkan pada Tabel 6.11 hanya 12 periode sedangkan dalam pembentukan model terdapat 108 periode yaitu dari tahun 2010 sampai tahun 2018.

6.1.6 Hasil Normalisasi Data

Normalisasi dilakukan dengan menggunakan normasilasi min max dengan fungsi ‘MinMaxScaler’. Hasil normalisasi yang ditunjukkan sebagai contoh hanya 12 periode pada tahun 2010.

Salah satu hasil normalisasi pada kecamatan Ngajum di dataran tinggi dengan skenario Top 3 IV ditunjukkan pada Tabel 6.12.

Tabel 6.12.Hasil Normalisasi Top 3 IV di Kecamatan Ngajum

Periode	Suhu udara	CH	JP
Jan-2010	0.567633	0.396383	0
Feb-2010	0.640097	0.453655	0.010747
Mar-2010	0.657005	0.325923	0.021502
Apr-2010	0.63285	0.390731	0.032265
Mei-2010	0.700483	0.27694	0.043036
Jun-2010	0.471014	0.347777	0.053814
Jul-2010	0.405797	0.081387	0.064601
Agt-2010	0.405797	0.148455	0.075396
Sep-2010	0.497585	0.354182	0.086199
Okt-2010	0.608696	0.457046	0.097011
Nop-2010	0.562802	0.660512	0.10783
Des-2010	0.521739	0.318387	0.118657

Untuk hasil normalisasi *dependent variable* yaitu jumlah KDB ditunjukkan pada Tabel 6.13.

6.1.7 Hasil Pembentukan Lag

Pembentukan dilakukan dengan melibatkan *dependent variable* yaitu jumlah KDB. Penambahan kolom dan berkurangnya baris pada data berdasarkan jumlah lag yang dibentuk. Hasil dari pembentukn lag ini digunakan untuk input atau *independent variable* dalam pembentukan model. Salah satu contoh hasil pembentukan lag pada kecamatan Ngajum di dataran tinggi dengan skenario Top 3 IV ditunjunjukkan pada Tabel 6.14.

Pada Tabel 6.14 dapat diketahui bahwa 2 periode yang berupa baris menghilang dan terdapat penambahan 2 kolom sesuai dengan lag yaitu kolom KDB_1 dan KDB_2. Jumlah lag yang dilakukan pada percobaan yaitu lag 0 sampai lag 12. Lag diterapkan pada semua skenario baik skenario pembagian data maupun skenario *independent variable*. Selain itu, jumlah dependent variable yang digunakan saat pembentukan model juga harus sama dengan jumlah *independent variable*. Jika

jumlah lag 2 maka pada *dependent variable* juga menghilangkan jumlah baris atau periode. Tabel 6.14 merupakan salah satu contoh input yang digunakan saat melakukan pembentukan model.

Tabel 6.13. Hasil Normalisasi Dependent Variable di Kecamatan Ngajum

Periode	KDB	Periode	KDB
Jan-2010	0.035714	Jul-2010	0
Feb-2010	0.107143	Agt-2010	0
Mar-2010	0	Sep-2010	0
Apr-2010	0.035714	Okt-2010	0
Mei-2010	0	Nop-2010	0
Jun-2010	0	Des-2010	0.035714

Tabel 6.14. Hasil Pembentukan Lag Skenario Top 3 IV di Kecamatan Ngajum

Periode	Suhu	CH	JP	KDB_1	KDB_2
Mar-2010	0.657005	0.325923	0.021502	0.107143	0.035714
Apr-2010	0.63285	0.390731	0.032265	0	0.107143
Mei-2010	0.700483	0.27694	0.043036	0.035714	0
Jun-2010	0.471014	0.347777	0.053814	0	0.035714
Jul-2010	0.405797	0.081387	0.064601	0	0
Agt-2010	0.405797	0.148455	0.075396	0	0
Sep-2010	0.497585	0.354182	0.086199	0	0
Okt-2010	0.608696	0.457046	0.097011	0	0
Nop-2010	0.562802	0.660512	0.10783	0	0
Des-2010	0.521739	0.318387	0.118657	0	0

6.2 Hasil Pembentukan Model pada Dataran Tinggi

Pada dataran tinggi, pembentukan model dilakukan pada kecamatan Ngajum sebagai kecamatan pembentuk model. Salah satu contoh hasil dari pembentukan model dengan skenario pembagian 80% data *training* dan 20% data *testing* ditunjukkan pada Tabel 6.15.

Tabel 6.15. Hasil Pembentukan Model dengan Pembagian Data 80% : 20%

Skenario Pembagian Data	Skenario IV	Lag	n_ estimator	Min_ Samples_ Split	Max_ Depth	RMSE Testing
80% : 20%	All IV	12	300	8	10	1.4921
	Top 3 IV	12	300	4	10	1.4923
	Top 2 IV	0	700	12	10	1.5054
	Top 1 IV	12	600	4	10	1.6004
	No IV	12	1000	8	10	1.7793

Pada Tabel 6.15 menunjukkan bahwa nilai RMSE terkecil sebesar 1.4921. Model terbaik yang didapatkan dari pembentukan model dengan pembagian 80% data *training* dan 20% data *testing* berada pada skenario *All Variable* dengan Lag 12. Parameter yang dipilih dari model terbaik yaitu jumlah n_estimator sebanyak 300, jumlah min_samples_split sebanyak 8, dan jumlah max_depth sebanyak 10.

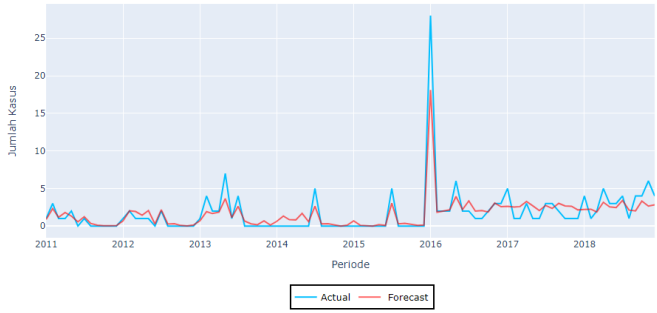
Namun, hasil model terbaik harus dibandingkan dengan hasil model terbaik dengan pembentukan model dengan skenario pembagian data yang lain. Hasil model dari skenario pembagian data yang lain dapat dilihat di Lampiran C. Untuk rangkuman hasil model terbaik dari pembentukan model setiap pembagian data yang berbeda ditunjukkan pada Tabel 6.16.

Tabel 6.16. Hasil Rangkuman Model Terbaik dari Setiap Skenario Pembagian Data di Dataran Tinggi

Skenario Pembagian Data	Skenario IV	Lag	n_ estimator	Min_ Samples_ Split	Max_ Depth	RMSE Testing
80% :20%	All IV	12	300	8	10	1.4921
70% : 30%	Top 3 IV	12	1000	12	10	1.3437
60% : 40%	Top 2 IV	6	200	16	5	4.3189
50% : 50%	Top 2 IV	4	400	12	5	3.9955

Pada Tabel 6.16 dapat diketahui bahwa nilai RMSE terkecil sebesar 1.3437 dengan pembagian 70% data *training* dan 30% data *testing*. Model terbaik yang dihasilkan di dataran tinggi berada pada skenario Top 3 IV dengan jumlah lag 12 dengan parameter jumlah n_estimator sebanyak 1000, jumlah

min_samples_split sebanyak 12, dan jumlah max_depth sebanyak 10. Model yang dihasilkan memperhatikan *independent variable* curah hujan, jumlah penduduk, suhu udara, dan variabel lag 12. Grafik perbandingan data aktual dengan hasil prediksi pada *data testing* ditunjukkan pada Gambar .



Gambar 6.2. Perbandingan data aktual dengan hasil prediksi pada data testing kecamatan Ngajum

Pada Gambar 6.2. terlihat bahwa model terbaik jika diujikan terhadap *data testing* dapat mengikuti pola data aktual. Namun, pada periode 2017 sampai 2018 hasil prediksi tidak cukup mengikuti data aktual dimana saat data aktual jumlah KDB turun, pada hasil prediksi naik. Sedangkan saat data aktual jumlah KDB naik, hasil prediksi menurun yang mana terlihat sekali pada akhir periode 2018.

6.2.1. Validasi Model terhadap Kecamatan Lain di Dataran Tinggi

Proses validasi model terbaik di dataran tinggi hanya menggunakan model terbaik yang dihasilkan dari setiap skenario pembentukan model pada kecamatan Ngajum. Sedangkan untuk melihat model yang bersifat *robust* menggunakan semua model hasil dari pembentukan model di setiap skenario pembagian data.

Model terbaik yang digunakan untuk validasi di dataran tinggi terhadap kecamatan lain yaitu skenario Top 3 IV lag 12 dengan

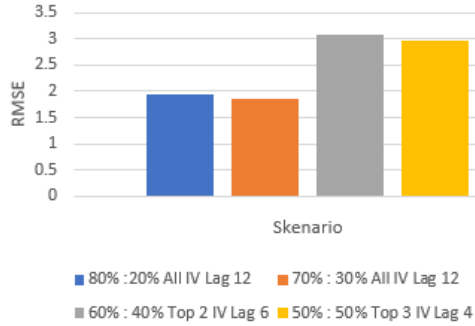
pembagian 70% data *training* dan 30% data *testing*. Hasil validasi model terbaik di dataran tinggi ditunjukkan pada Tabel 6.17.

Tabel 6.17. Hasil Validasi Kecamatan dari Model Terbaik di Dataran Tinggi

Kecamatan	RMSE	SMAPE
Ngajum	1.3437	24.1388
Poncokusumo	2.2031	43.7872
Jabung	2.0711	39.0961

Pada Tabel 6.17 dapat dilihat bahwa model terbaik pada kecamatan Poncokusumo memiliki nilai RMSE sebesar 2.2031 dan nilai SMAPE sebesar 43.79%. Sedangkan pada kecamatan Jabung memiliki nilai RMSE sebesar 2.0711 dan nilai SMAPE 39.10%.

Model terbaik dari dataran tinggi merupakan model yang tidak *robust* terhadap validasi kecamatan lain karena rata-rata nilai RMSE terkecil berada pada model lain. Untuk hasil rangkuman model dengan rata-rata nilai RMSE terkecil dari setiap skenario pembagian data ditunjukkan pada Gambar 6.3.



Gambar 6.3. Hasil Rangkuman Validasi Kecamatan dari Semua Skenario di Dataran Tinggi

Pada Gambar 6.3 dapat dilihat bahwa nilai rata-rata RMSE terkecil sebesar 1.8646. Model *robust* pada dataran tinggi berada pada pembagian 70% data *training* dan 30% data *testing*

dengan skenario All IV lag 12. Model *robust* terhadap kecamatan lain di dataran tinggi memperhatikan semua *independent variable*. Untuk hasil semua validasi kecamatan di semua skenario dapat dilihat di Lampiran D.

6.2.2. Validasi Model terhadap Pembagian Data di Dataran Tinggi

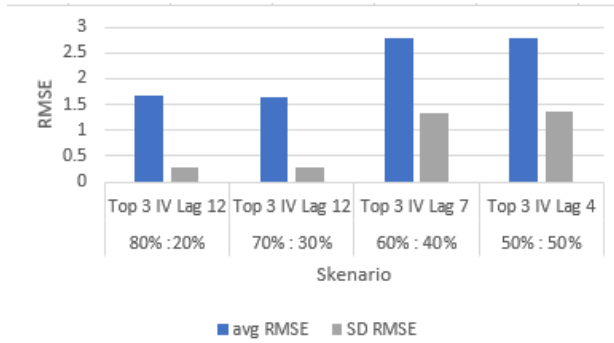
Validasi model terhadap pembagian data di dataran tinggi menggunakan model terbaik pada kecamatan Ngajum yaitu skenario Top 3 IV lag 12 dengan pembagian 70% data *training* dan 30% data *testing*. Untuk hasil validasi model terbaik terhadap pembagian data ditunjukkan pada Tabel 6.18.

Tabel 6.18. Hasil Validasi Pembagian Data dari Model Terbaik di Dataran Tinggi

Validasi Pembagian Data	RMSE	SMAPE
80% : 20%	1.4043	24.8487
60% : 40%	2.0033	27.6973
50% : 50%	1.8322	39.6809

Pada tabel 6.18. menunjukkan bahwa nilai RMSE terkecil berada di validasi pembagian 80% data *training* dan 20% data *testing* dengan nilai RMSE sebesar 1.4043 dan nilai SMAPE sebesar 24.85%. Model terbaik di dataran tinggi merupakan model yang *robust* terhadap pembagian data karena hasil nilai rata-rata dan standard deviasi RMSE terkecil juga berada pada model terbaik yang telah dihasilkan.

Hasil rangkuman model dengan rata-rata dan standar deviasi nilai RMSE terkecil dari proses validasi pembagian data dari model terbaik semua skenario ditunjukkan pada Gambar 6.4. Jika dilihat dari rata-rata dan standar deviasi nilai RMSE pada semua validasi pembagian data model *robust* berada pada skenario Top 3 IV lag 12 dengan pembagian 70% data *training* dan 30% data *testing.* Untuk hasil semua model validasi pembagian data dari semua skenario dapat dilihat pada Lampiran D.



Gambar 6.4. Hasil Rangkuman Validasi Pembagian Data dari Model Terbaik Semua Skenario di Dataran Tinggi

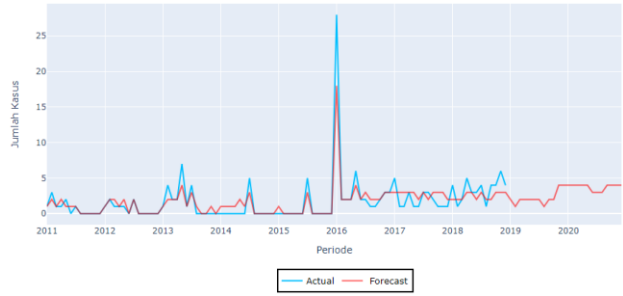
6.2.3. Hasil Peramalan 24 Periode Mendatang di Dataran Tinggi

Pada dataran tinggi, model terbaik yang digunakan yaitu model skenario Top 3 IV lag 12 dari pembentukan model 70% data *training* dan 30% data *testing* pada kecamatan Ngajum. Model ini memperhatikan *independent variable* curah hujan, jumlah penduduk, suhu udara dan variabel lag 12.

a. Kecamatan Ngajum

Hasil peramalan untuk 24 periode mendatang jumlah KDB pada kecamatan Ngajum ditunjukkan pada Tabel 6.19.

Grafik perbandingan data aktual dan peramalan jumlah KDB pada kecamatan Ngajum ditunjukkan pada Gambar 6.5.



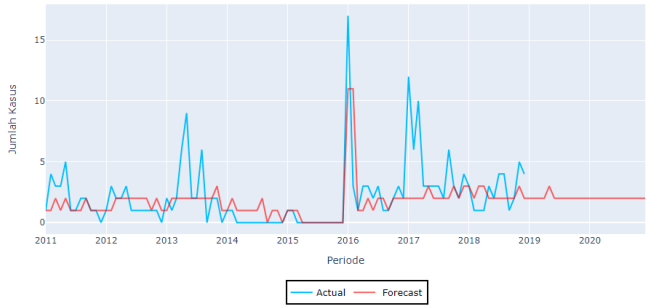
Gambar 6.5. Hasil Peramalan pada Kecamatan Ngajum

Tabel 6.19.Hasil Peramalan 24 Periode Mendatang pada Kecamatan Ngajum

Periode	Jumlah KDB	Periode	Jumlah KDB
Jan-2019	2	Jan-2020	4
Feb-2019	1	Feb-2020	4
Mar-2019	2	Mar-2020	4
Apr-2019	2	Apr-2020	4
Mei-2019	2	Mei-2020	4
Jun-2019	2	Jun-2020	3
Jul-2019	2	Jul-2020	3
Agt-2019	1	Agt-2020	3
Sep-2019	2	Sep-2020	4
Okt-2019	2	Okt-2020	4
Nop-2019	4	Nop-2020	4
Des-2019	4	Des-2020	4

b. Kecamatan Poncokusumo

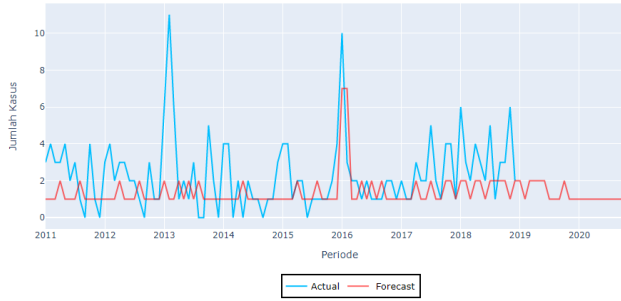
Grafik perbandingan data aktual dan peramalan jumlah KDB pada kecamatan Poncokusumo ditunjukkan pada Gambar 6.6. Untuk hasil peramalan 24 periode mendatang pada kecamatan Poncokusumo dapat dilihat pada Lampiran E.



Gambar 6.6. Hasil Peramalan pada Kecamatan Poncokusumo

c. Kecamatan Jabung

Grafik perbandingan data aktual dan peramalan jumlah KDB pada kecamatan Jabung ditunjukkan pada Gambar 6.7. Untuk hasil peramalan 24 periode mendatang pada kecamatan Jabung dapat dilihat pada Lampiran E.



Gambar 6.7.Hasil Peramalan pada Kecamatan Jabung

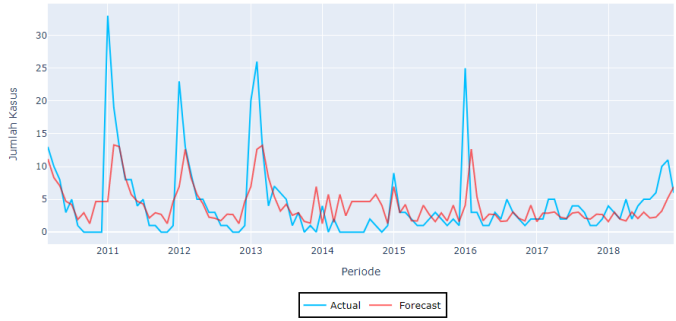
6.3 Hasil Pembentukan Model pada Dataran Sedang

Pada dataran sedang, pembentukan model dilakukan di kecamatan Pakishaji sebagai kecamatan pembentuk model. Hasil dari pembentukan model di setiap skenario pembagian data dapat dilihat pada Lampiran C. Untuk rangkuman hasil model terbaik dari setiap skenario pembagian data ditunjukkan pada Tabel 6.20.

Pada Tabel 6.20 dapat dilihat bahwa nilai RMSE terkecil sebesar 2.5964 dengan pembagian 80% data *training* dan 20% data *testing*. Model terbaik yang dihasilkan di dataran tinggi berada pada skenario No IV dengan jumlah lag 2 dengan parameter jumlah n_estimator sebanyak 200, jumlah min_samples_split sebanyak 16, dan jumlah max_depth sebanyak 10. Model ini hanya memperhatikan variabel lag sebanyak 2. Grafik perbandingan data aktual dengan hasil prediksi pada *data testing* ditunjukkan pada Gambar 6.8.

Tabel 6.20. Hasil Rangkuman Model Terbaik dari Setiap Skenario Pembagian Data di Dataran Sedang

Skenario Pembagian Data	Skenario IV	Lag	n_ estimator	Min_ Samples_ Split	Max_ Depth	RMSE Testing
80% :20%	No IV	2	200	16	10	2.5964
70% : 30%	Top 1	3	100	4	5	2.7650
60% : 40%	Top 1	2	100	12	5	4.4976
50% : 50%	Top 2	3	400	12	5	4.4052



Gambar 6.8. Perbandingan data aktual dengan hasil prediksi pada data testing kecamatan Pakishaji

Pada Gambar 6.8. terlihat bahwa model terbaik jika diujikan terhadap *data testing* cukup dapat mengikuti pola data aktual. Namun, pada periode 2014 sampai 2015 hasil prediksi cenderung meningkat dibandingkan dengan data aktual. Untuk periode 2016 sampai 2018 hasil prediksi cenderung mengalami keterlambatan untuk mengikuti pola dari data aktual.

6.3.1. Validasi Model terhadap Kecamatan Lain di Dataran Sedang

Model terbaik yang digunakan untuk validasi di dataran sedang terhadap kecamatan lain yaitu skenario No IV lag 2 dengan pembagian 80% data *training* dan 20% data *testing*. Hasil validasi model terbaik ditunjukkan pada Tabel 6.21.

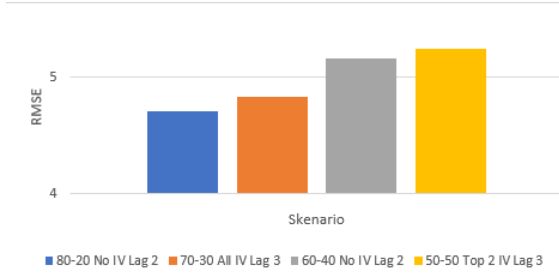
Pada Tabel 6.21 dapat diketahui bahwa model terbaik yang dihasilkan dari kecamatan Pakishaji paling cocok diterapkan pada kecamatan Wajak, Singosari dan Karangploso. Jika dilihat

dari nilai RMSE, kecamatan Wajak, Singosari dan Karangploso memiliki nilai RMSE terkecil yaitu 3.4700, 3.1360 dan 3.1700 sedangkan nilai SMAPE sebesar 47,74%, 51.87% dan 39.69%.

Tabel 6.21. Hasil Validasi Kecamatan dari Model Terbaik di Dataran Sedang

Kecamatan	RMSE	SMAPE
Pakishaji	2.5965	35.5511
Dampit	4.5618	47.3052
Tumpang	4.7728	34.9379
Wajak	3.4700	47.7427
Singosari	3.1360	51.8703
Sumbermanjing	8.9158	59.9202
Lawang	7.0548	42.8204
Karangploso	3.1700	39.6892

Model terbaik dari dataran sedang merupakan model yang *robust* terhadap validasi kecamatan lain karena rata-rata nilai RMSE terkecil juga berada pada model terbaik yang telah dihasilkan. Nilai rata-rata RMSE pada model terbaik sebesar 4.7097. Hasil rangkuman model dengan rata-rata nilai RMSE terkecil dari setiap skenario pembagian data ditunjukkan pada Gambar 6.9.



Gambar 6.9. Hasil Rangkuman Validasi Kecamatan dari Semua Skenario di Dataran Sedang

Untuk hasil validasi kecamatan semua model dengan rata-rata nilai RMSE terkecil dari setiap skenario pembagian data dapat dilihat pada Lampiran D.

6.3.2. Validasi Model terhadap Pembagian Data di Dataran Sedang

Validasi model terhadap pembagian data di dataran sedang menggunakan model terbaik pada kecamatan Pakishaji yaitu skenario No IV lag 2 dengan pembagian 80% data *training* dan 20% data *testing*. Untuk hasil validasi model terbaik terhadap pembagian data ditunjukkan pada Tabel 6.22.

Tabel 6.22. Hasil Validasi Pembagian Data dari Model Terbaik di Dataran Sedang

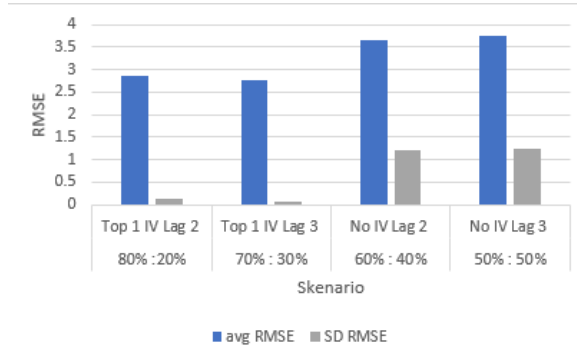
Validasi Pembagian Data	RMSE	SMAPE
70% : 30%	2.6142	34.8951
60% : 40%	4.6953	34.7714
50% : 50%	4.4769	34.4741

Pada Tabel 6.22 menunjukkan bahwa nilai RMSE terkecil berada di validasi pembagian 70% data training dan 30% data testing dengan nilai RMSE sebesar 1.4043 sedangkan nilai SMAPE sebesar 34.89%. Model terbaik di dataran sedang merupakan model yang tidak *robust* terhadap pembagian data karena hasil nilai rata-rata dan standar devisiasi RMSE yang terkecil berada pada model yang lain.

Hasil rangkuman model dengan rata-rata dan standar deviasi nilai RMSE terkecil dari proses validasi pembagian data dari model terbaik semua skenario ditunjukkan pada Gambar 6.10.

Pada Gambar 6.10. dapat diketahui bahwa nilai rata-rata RMSE terkecil sebesar 2.7773 dan standar deviasi sebesar 0.0570. Model *robust* pada dataran sedang berada pada skenario Top 1 IV lag 3 dengan pembagian data sebesar 70% data *training* dan 30% data *testing*. *Independent variable* yang memengaruhi yaitu variabel angka bebas jentik. Untuk hasil semua model

validasi pembagian data dari semua skenario dapat dilihat pada Lampiran D.



Gambar 6.10. Hasil Rangkuman Validasi Pembagian Data dari Model Terbaik Semua Skenario di Dataran Sedang

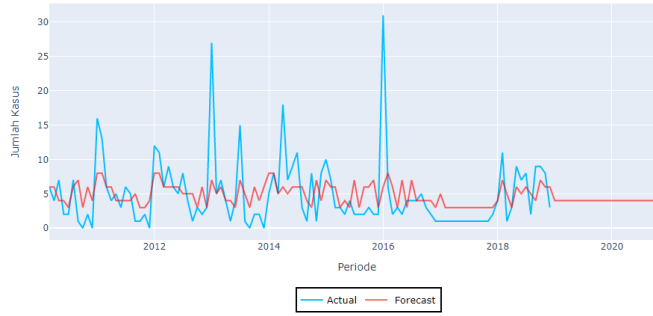
6.3.3. Hasil Peramalan 24 Periode Mendatang di Dataran Sedang

Pada dataran sedang, model terbaik yang digunakan yaitu model skenario No IV lag 2 dari pembentukan model 80% data *training* dan 20% data *testing* pada kecamatan Pakishaji.

a. Kecamatan Pakishaji

Hasil peramalan untuk 24 periode mendatang jumlah KDB pada kecamatan Pakishaji ditunjukkan pada Tabel 6.23.

Grafik perbandingan data aktual dan peramalan jumlah KDB pada kecamatan Pakishaji ditunjukkan pada Gambar 6.11.



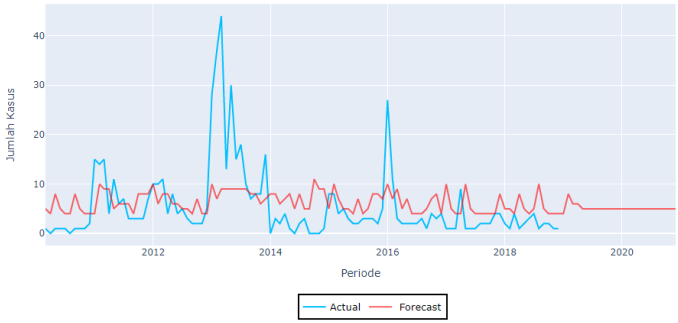
Gambar 6.11. Hasil Peramalan pada Kecamatan Pakishaji

Tabel 6.23. Hasil Peramalan 24 Periode Mendatang pada Kecamatan Pakishaji

Periode	Jumlah KDB	Periode	Jumlah KDB
Jan-2019	4	Jan-2020	4
Feb-2019	4	Feb-2020	4
Mar-2019	4	Mar-2020	4
Apr-2019	4	Apr-2020	4
Mei-2019	4	Mei-2020	4
Jun-2019	4	Jun-2020	4
Jul-2019	4	Jul-2020	4
Agt-2019	4	Agt-2020	4
Sep-2019	4	Sep-2020	4
Okt-2019	4	Okt-2020	4
Nop-2019	4	Nop-2020	4
Des-2019	4	Des-2020	4

a. Kecamatan Lawang

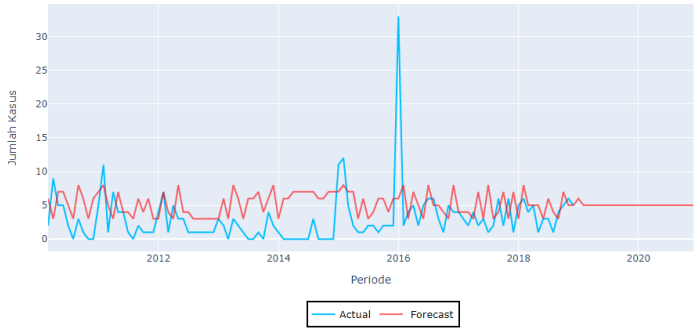
Grafik perbandingan data aktual dan peramalan jumlah KDB pada kecamatan Lawang ditunjukkan pada Gambar 6.12. Untuk hasil peramalan 24 periode mendatang pada kecamatan Lawang dapat dilihat pada Lampiran E.



Gambar 6.12. Hasil Peramalan pada Kecamatan Lawang

a. Kecamatan Dampit

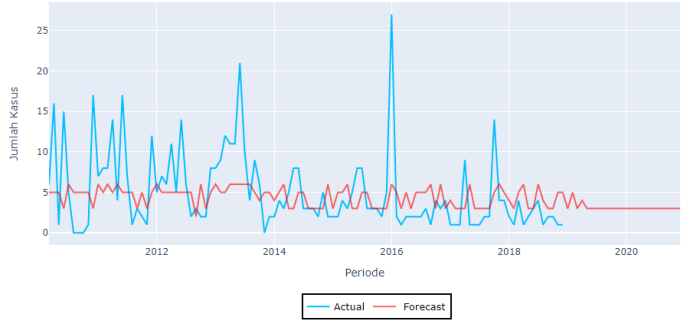
Grafik perbandingan data aktual dan peramalan jumlah KDB pada kecamatan Dampit ditunjukkan pada Gambar 6.13 Untuk hasil peramalan 24 periode mendatang pada kecamatan Dampit dapat dilihat pada Lampiran E.



Gambar 6.13. Hasil Peramalan pada Kecamatan Dampit

b. Kecamatan Tumpang

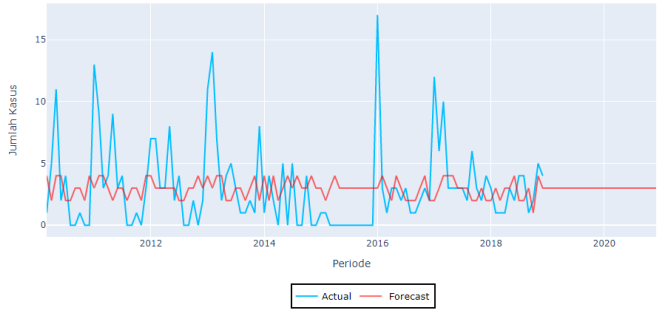
Grafik perbandingan data aktual dan peramalan jumlah KDB pada kecamatan Tumpang ditunjukkan pada Gambar 6.14. Untuk hasil peramalan 24 periode mendatang pada kecamatan Tumpang dapat dilihat pada Lampiran E.



Gambar 6.14. Hasil Peramalan pada Kecamatan Tumpang

c. Kecamatan Wajak

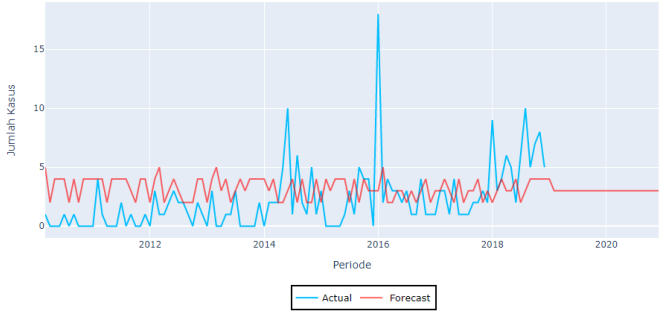
Grafik perbandingan data aktual dan peramalan jumlah KDB pada kecamatan Wajak ditunjukkan pada Gambar 6.15. Untuk hasil peramalan 24 periode mendatang pada kecamatan Wajak dapat dilihat pada Lampiran E.



Gambar 6.15. Hasil Peramalan pada Kecamatan Wajak

d. Kecamatan Singosari

Grafik perbandingan data aktual dan peramalan jumlah KDB pada kecamatan Singosari ditunjukkan pada Gambar 6.16. Untuk hasil peramalan 24 periode mendatang pada kecamatan Singosari dapat dilihat pada Lampiran E.

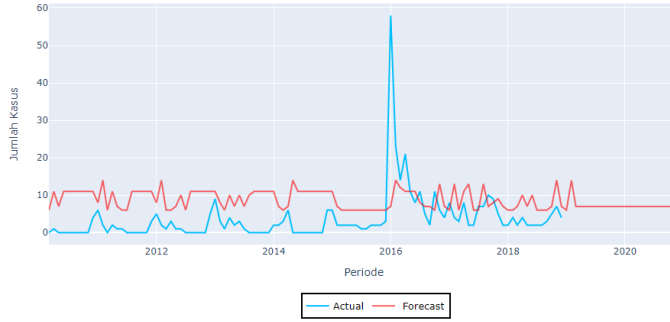


Gambar 6.16. Hasil Peramalan pada Kecamatan Singosari

e. Kecamatan Sumbermanjing

Grafik perbandingan data aktual dan peramalan jumlah KDB pada kecamatan Sumbermanjing ditunjukkan pada Gambar

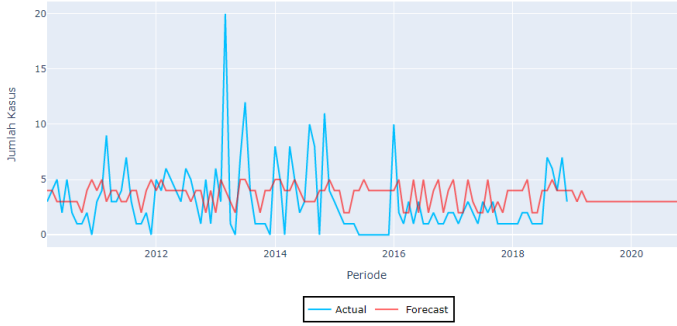
6.17. Untuk hasil peramalan 24 periode mendatang pada kecamatan Sumbermanjing dapat dilihat pada Lampiran E.



Gambar 6.17. Hasil Peramalan pada Kecamatan Sumbermanjing

f. Kecamatan Karangploso

Grafik perbandingan data aktual dan peramalan jumlah KDB pada kecamatan Karangploso ditunjukkan pada Gambar 6.18. Untuk hasil peramalan 24 periode mendatang pada kecamatan Karangploso dapat dilihat pada Lampiran E.



Gambar 6.18. Hasil Peramalan pada Kecamatan Karangploso

6.4 Hasil Pembentukan Model pada Dataran Rendah

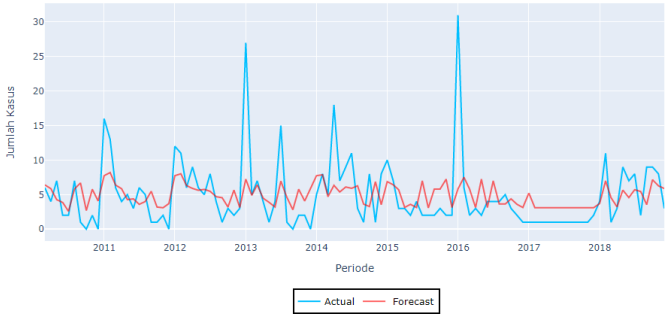
Pada dataran rendah, pembentukan model dilakukan di kecamatan Kepanjen sebagai kecamatan pembentuk model. Hasil dari pembentukan model di setiap skenario pembagian data dapat dilihat pada Lampiran C. Untuk rangkuman hasil

model terbaik dari setiap skenario pembagian data ditunjukkan pada Tabel 6.24.

Tabel 6.24. Hasil Rangkuman Model Terbaik dari Setiap Skenario Pembagian Data di Dataran Rendah

Skenario Pembagian Data	Skenario IV	Lag	n_ estimator	Min_ Samples_ Split	Max_ Depth	RMSE Testing
80% :20%	Top 1	2	200	16	10	2.3046
70% : 30%	No IV	2	700	16	5	2.2568
60% : 40%	Top 1	0	900	4	15	3.9256
50% : 50%	No IV	1	400	8	5	3.9496

Pada Tabel 6.24 menunjukkan bahwa nilai RMSE terkecil sebesar 2.2568 terdapat pada pembagian 70% data *training* dan 30% data *testing*. Model terbaik yang dihasilkan di dataran tinggi berada pada skenario No IV dengan jumlah lag 2 dengan parameter jumlah n_estimator sebanyak 700, jumlah min_samples_split sebanyak 16, dan jumlah max_depth sebanyak 5. Model ini hanya memperhatikan variabel lag sebanyak 2. Grafik perbandingan data aktual dengan hasil prediksi pada *data testing* ditunjukkan pada Gambar 6.19.



Gambar 6.19. Perbandingan data aktual dengan hasil prediksi pada data testing kecamatan Kepanjen

Pada Gambar 6.19 terlihat bahwa model terbaik jika diujikan terhadap *data testing* cukup dapat mengikuti pola data aktual. Namun, pada periode pertengahan 2015 dan 2017, hasil prediksi

cenderung meningkat dibandingkan dengan data aktual dimana data aktual menurun sedangkan hasil prediksi naik.

Jika dilihat dari periode 2011 sampai 2018, hasil prediksi cenderung mengalami keterlambatan dalam mengikuti pola data aktual karena terdapat jumlah KDB yang cukup tinggi di beberapa periode.

6.4.1. Validasi Model terhadap Kecamatan Lain di Dataran Rendah

Proses validasi model terbaik di dataran rendah dilakukan pada model terbaik yang dihasilkan dari setiap skenario pembentukan model. Sedangkan untuk melihat model yang bersifat *robust* menggunakan semua model hasil dari pembentukan model di setiap skenario pembagian data pada kecamatan Kepanjen.

Model terbaik yang digunakan untuk validasi di dataran rendah terhadap kecamatan lain yaitu skenario No IV lag 2 dengan pembagian 70% data *training* dan 30% data *testing*. Hasil validasi model terbaik di dataran rendah ditunjukkan pada Tabel 6.25.

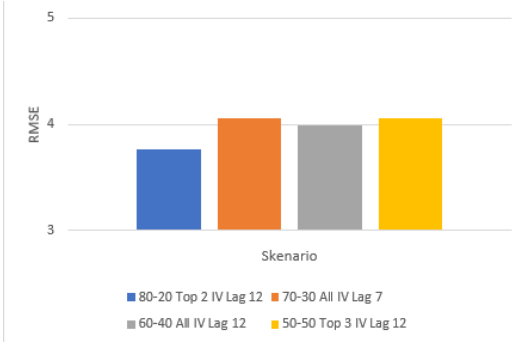
Tabel 6.25. Hasil Validasi Kecamatan dari Model Terbaik di Dataran Rendah

Kecamatan	RMSE	SMAPE
Kepanjen	2.2569	23.5219
Turen	14.9031	39.4672
Gondanglegi	2.3164	48.8334
Donomulyo	2.5560	53.7126
Bululawang	3.0545	32.6746

Pada tabel 6.25 dapat diketahui bahwa model terbaik yang dihasilkan dari kecamatan Kepanjen paling cocok diterapkan pada kecamatan Gondanglegi, Donomulyo, dan Bululawang. Jika dilihat dari nilai RMSE, kecamatan Gondanglegi, Donomulyo, dan Bululawang memiliki nilai RMSE terkecil yaitu 2.3164, 2.5560 dan 3.0545 sedangkan nilai SMAPE sebesar 48.83%, 53.71% dan 32.67%. Pada kecamatan Turen

nilai RMSE adalah nilai yang tinggi sebesar 14.9031 karena data yang digunakan dalam pembentukan model terdapat *outbreak*.

Model terbaik dari dataran rendah merupakan model yang tidak *robust* terhadap validasi kecamatan lain karena rata-rata nilai RMSE terkecil berada pada model yang lain. Hasil rangkuman model dengan rata-rata nilai RMSE terkecil dari setiap skenario pembagian data ditunjukkan pada Gambar 6.20.



Gambar 6.20. Hasil Rangkuman Validasi Kecamatan dari Semua Skenario di Dataran Rendah

Jika dilihat dari rata-rata nilai RMSE pada semua kecamatan, model *robust* berada pada skenario Top 2 IV lag 12 dengan pembagian 80% data *training* dan 20% data *testing* yang memiliki nilai rata-rata RMSE sebesar 3.7590. *Independent variable* yang memengaruhi yaitu variabel jumlah penduduk dan kelembapan udara. Untuk hasil validasi kecamatan semua model dengan rata-rata nilai RMSE terkecil dari setiap skenario pembagian data dapat dilihat pada Lampiran D.

6.4.2. Validasi Model terhadap Pembagian Data di Dataran Rendah

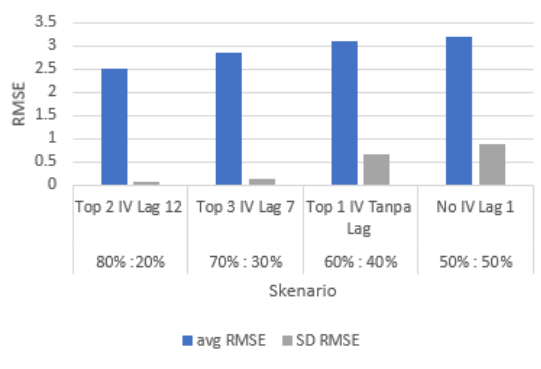
Validasi model terhadap pembagian data di dataran rendah menggunakan model terbaik pada kecamatan Kepanjen yaitu skenario No IV lag 2 dengan pembagian 70% data *training* dan 30% data *testing*. Untuk hasil validasi model terbaik terhadap pembagian data ditunjukkan pada Tabel 6.26.

Tabel 6.26. Hasil Validasi Pembagian Data dari Model Terbaik di Dataran Rendah

Validasi Pembagian Data	RMSE	SMAPE
80% : 20%	2.5504	24.9124
60% : 40%	4.0847	26.7378
50% : 50%	3.8654	28.9962

Pada tabel 6.26 menunjukkan bahwa nilai RMSE terkecil berada di validasi pembagian 80% data *training* dan 20% data *testing* dengan nilai RMSE sebesar 2.5504 sedangkan nilai SMAPE sebesar 24.91%. Model terbaik di dataran rendah merupakan model yang tidak *robust* terhadap pembagian data karena hasil nilai rata-rata dan standard deviasi RMSE yang terkecil berada pada model yang lain.

Hasil rangkuman model dengan rata-rata dan standar deviasi nilai RMSE terkecil dari proses validasi pembagian data dari model terbaik semua skenario ditunjukkan pada Gambar 6.21.



Gambar 6.21. Hasil Rangkuman Validasi Pembagian Data dari Model Terbaik Semua Skenario di Dataran Rendah

Pada Gambar 6.21 dapat diketahui bahwa nilai rata-rata RMSE terkecil sebesar 2.5151 dan standar deviasi sebesar 0.0877. Model *robust* pada dataran rendah berada pada skenario Top 2 IV lag 12 dengan pembagian data sebesar 80% data *training* dan 20% data *testing*. Model ini memperhatikan *Independent*

variable jumlah penduduk dan kelembapan udara. Untuk hasil semua model validasi pembagian data dari semua skenario dapat dilihat pada Lampiran D.

6.4.3. Hasil Peramalan 24 Periode Mendatang di Dataran Rendah

Pada dataran sedang, model terbaik yang digunakan yaitu model skenario No IV lag 2 dari pembentukan model 70% data *training* dan 30% data *testing* pada kecamatan Kepanjen.

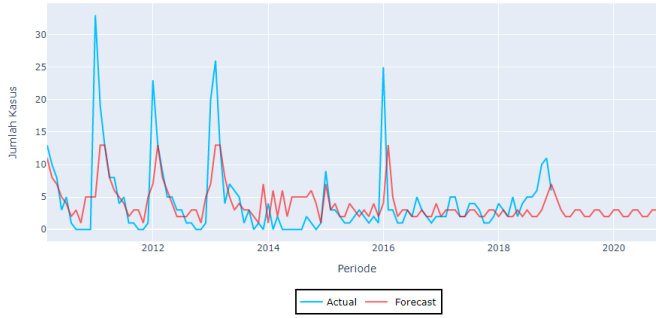
a. Kecamatan Kepanjen

Hasil peramalan untuk 24 periode mendatang jumlah KDB pada kecamatan Kepanjen ditunjukkan pada Tabel 6.27.

Tabel 6.27. Hasil Peramalan 24 Periode Mendatang pada Kecamatan Kepanjen

Periode	Jumlah KDB	Periode	Jumlah KDB
Jan-2019	5	Jan-2020	3
Feb-2019	3	Feb-2020	3
Mar-2019	2	Mar-2020	2
Apr-2019	2	Apr-2020	2
Mei-2019	3	Mei-2020	3
Jun-2019	3	Jun-2020	3
Jul-2019	2	Jul-2020	2
Agt-2019	2	Agt-2020	2
Sep-2019	3	Sep-2020	3
Okt-2019	3	Okt-2020	3
Nop-2019	2	Nop-2020	2
Des-2019	2	Des-2020	2

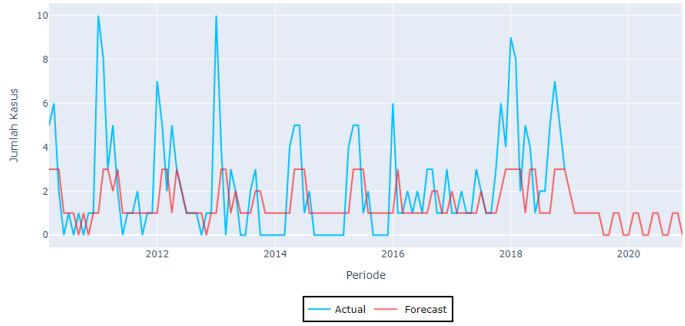
Grafik perbandingan data aktual dan peramalan jumlah KDB pada kecamatan Kepanjen ditunjukkan pada Gambar 6.22.



Gambar 6.22. Hasil Peramalan pada Kecamatan Kepanjen

b. Kecamatan Gondanglegi

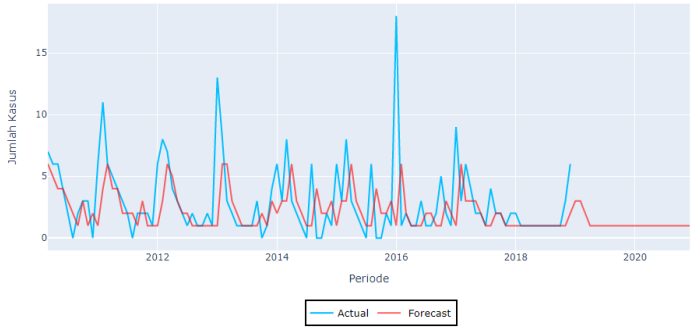
Grafik perbandingan data aktual dan peramalan jumlah KDB pada kecamatan Gondanglegi ditunjukkan pada Gambar 6.23. Untuk hasil peramalan 24 periode mendatang pada kecamatan Gondanglegi dapat dilihat pada Lampiran E.



Gambar 6.23. Hasil Peramalan pada Kecamatan Gondanglegi

c. Kecamatan Bululawang

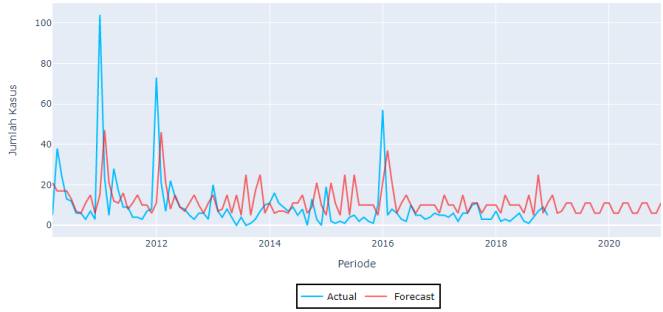
Grafik perbandingan data aktual dan peramalan jumlah KDB pada kecamatan Bululawang ditunjukkan pada Gambar 6.24. Untuk hasil peramalan 24 periode mendatang pada kecamatan Bululawang dapat dilihat pada Lampiran E.



Gambar 6.24. Hasil Peramalan pada Kecamatan Bululawang

d. Kecamatan Turen

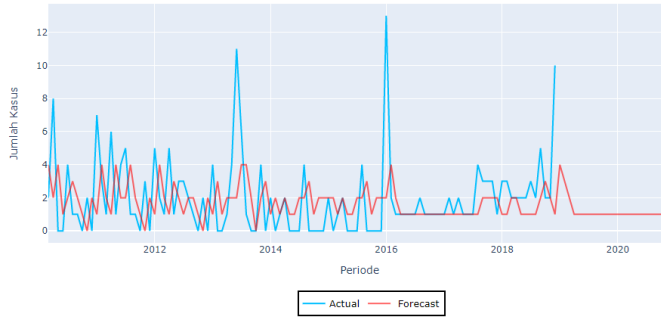
Grafik perbandingan data aktual dan peramalan jumlah KDB pada kecamatan Turen ditunjukkan pada Gambar 6.25. Untuk hasil peramalan 24 periode mendatang pada kecamatan Turen dapat dilihat pada Lampiran E.



Gambar 6.25. Hasil Peramalan pada Kecamatan Turen

e. Kecamatan Donomulyo

Grafik perbandingan data aktual dan peramalan jumlah KDB pada kecamatan Donomulyo ditunjukkan pada Gambar 6.26. Untuk hasil peramalan 24 periode mendatang pada kecamatan Donomulyo dapat dilihat pada Lampiran E.



Gambar 6.26. Hasil Peramalan pada Kecamatan Donomulyo

6.5 Hasil Pencarian Satu Model Terbaik Untuk Seluruh Kecamatan

Pada penilitian sebelumnya, terdapat 3 model terbaik yang digunakan untuk 3 dataran yang mana setiap dataran memiliki 1 model terbaik. Pembagian kecamatan di Kabupaten Malang menjadi 3 dataran karena setiap dataran mempunyai iklim yang berbeda-beda dimana iklim berpengaruh terhadap jumlah KDB. Hasil dari penelitian sebelumnya yaitu jumlah KDB di dataran tinggi lebih sedikit dibandingkan dengan dataran rendah dan dataran sedang. Hal ini yang menjadi dasar bahwa jumlah KDB dipengaruhi oleh iklim sehingga terdapat 3 model terbaik yang digunakan untuk melakukan peramalan di 3 dataran [9]. Sedangkan pada penelitian tugas akhir ini, pencarian 1 model terbaik dilakukan dengan tujuan model tersebut dapat digunakan untuk melakukan peramalan periode mendatang pada seluruh kecamatan di Kabupaten Malang.

Proses mencari model terbaik yang dapat diimplementasikan untuk seluruh kecamatan, pada penelitian tugas akhir ini menggunakan proporsi data yang dihasilkan dari model terbaik di setiap dataran. Setiap melakukan pembentukan model terdapat 5 skenario *independent variable*. Pada dataran tinggi dan rendah, hasil model terbaik didapatkan dari proporsi data 70% data training dan 30% data testing. Sedangkan pada dataran sedang, model terbaik dihasilkan dari pembentukan model 80% data training dan 20% data testing.

Setiap skenario *independent variable* di setiap pembentukan model dilakukan pengujian terhadap kecamatan lain di semua dataran. Jumlah dari semua model yang digunakan dalam pencarian satu model terbaik sebanyak 15 model. Pemilihan model terbaik untuk seluruh kecamatan berbeda dengan pemilihan model robust terhadap kecamatan lain di setiap dataran.

Pemilihan satu model terbaik untuk seluruh kecamatan berdasarkan median dari nilai RMSE *testing* pada seluruh kecamatan bukan berdasarkan rata-rata nilai RMSE *testing*. Penggunaan nilai median lebih bagus dibandingkan rata-rata nilai RMSE karena median tidak dipengaruhi oleh data esktrim. Sedangkan rata-rata sangat sensitif terhadap data ekstrim [27]. Pada penilitian ini terdapat di setiap kecamatan KDB yang memiliki nilai ekstrim ditandai dengan jumlah KDB yang meningkat bahkan terdapat *outbreak*. Untuk menghasilkan model yang dapat mewakili seluruh kecamatan maka pemilihan model terbaik untuk seluruh kecamatan menggunakan nilai median dari RMSE *testing*. Hasil pencarian satu model terbaik dari 15 model untuk seluruh kecamatan ditunjukkan pada Tabel 6.28.

Pada Tabel 6.28 dapat diketahui bahwa nilai median terkecil dari seluruh kecamatan sebesar 2.9675 yang didapatkan pada pembentukan model 70% data training dan 30% data testing di dataran tinggi. Skenario *independent variable* dari model terbaik untuk seluruh kecamatan yaitu All IV Lag 12. Model ini memperhatikan semua *independent variable* yaitu variabel suhu, curah hujan, kecepatan angin,kelembapan udara, angka bebas jentik, jumlah penduduk, dan variabel lag 12.

Tabel 6.28 Hasil pencarian satu model terbaik dari 15 model untuk seluruh kecamatan

Pembentukan Model	Skenario IV	Avg	Median
Dataran Tinggi [70% ; 30%]	All IV Lag 12	4.0258	2.9675
	Top 3 IV Lag 12	4.0511	3.0168
	Top 2 IV Lag 12	4.0399	3.0045
	Top 1 IV Lag 12	4.0127	2.9793
	No IV Lag 2	4.7913	3.7838
Dataran Sedang [80% ; 20%]	All IV Lag 2	4.7890	3.5104
	Top 3 IV Lag 3	5.5284	3.5987
	Top 2 IV Lag 2	5.0638	3.4123
	Top 1 IV Lag 2	5.5022	3.7574
	No IV Lag 2	5.0876	3.3200
Dataran Rendah [70% ; 30%]	No IV Lag 2	4.4063	3.3477
	All IV Lag 7	3.8750	3.0931
	Top 3 IV Lag 7	4.2209	3.1449
	Top 2 IV Lag 7	4.6797	3.2278
	Top 1 IV Lag 9	4.7792	3.3043

6.6 Hasil Perbandingan Metode

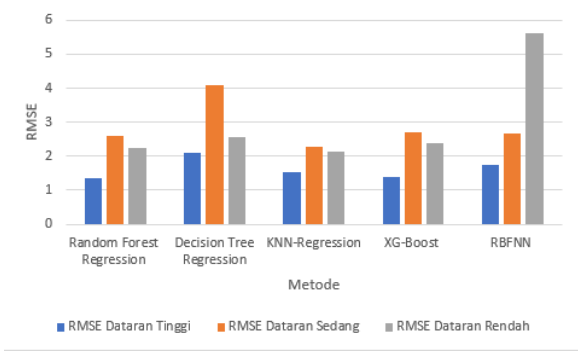
Pada sub bab ini menjelaskan penggunaan beberapa metode lain dengan tujuan untuk membandingkan hasil model terbaik di setiap dataran antara metode lain dengan metode *Random Forest Regression.* Penelitian sebelumnya yang menggunakan data jumlah kejadian demam berdarah di Kabupaten Malang sudah sekitar 11 dengan menggunakan metode yang berbeda diantaranya metode *Hybrid Autoregressive Integrated Moving Average - Neural Network Autoregressive with Exogenous* (ARIMA-NNARX), *Group Method Of Data Handling* (GMDH), *Particle Swarm Optimization - Extreme Learning Machine* (PSO-ELM), *Fuzzy Inference System, Backpropagation Neural Network, Radial Basis Function Neural Network* (RBFNN), dan *Autoregressive Integrated Moving Average Generelized Regression Neural Network*

(ARIMA-GRNN). Percobaan dengan menggunakan metode yang berbeda-beda bertujuan untuk membandingkan metode mana yang menghasilkan model terbaik untuk melakukan peramalan jumlah KDB di Kabupaten Malang pada periode mendatang.

Namun, metode yang digunakan sebagai pembanding dalam penelitian tugas akhir ini yaitu metode *KNN Regression, XG-Boost, Decision Tree Regression* , dan *Radial Basis Function Neural Network (RBFNN).* Perbandingan dilakukan berdasarkan pada nilai RMSE model terbaik pada kecamatan pembentuk model di setiap dataran. Pada perbandingan antar metode, terdapat metode yang tidak mengalami proses *tuning parameter* yaitu metode *Decision Tree Regression*. Sedangkan pada metode lain menggunakan proses *tuning parameter* untuk menghasilkan model dengan parameter terbaik. Hasil perbandingan metode lain dengan metode *Random Forest Regression* ditunjukkan pada Tabel 6.29. Untuk hasil visualisasi perbandingan RMSE ditunjukkan pada Gambar 6.27.

Tabel 6.29. Hasil Perbandingan Beberapa Metode

Metode	RMSE		
	Dataran Tinggi	**Dataran Sedang**	**Dataran Rendah**
Random Forest Regression	1.344	2.596	2.257
Decision Tree Regression	2.101	4.072	2.555
KNN-Regression	1.543	2.293	2.133
XG-Boost	1.385	2.699	2.399
RBFNN	1.753	2.685	5.623



Gambar 6.27. Perbandingan RMSE Random Forest Regression dengan Metode Lain

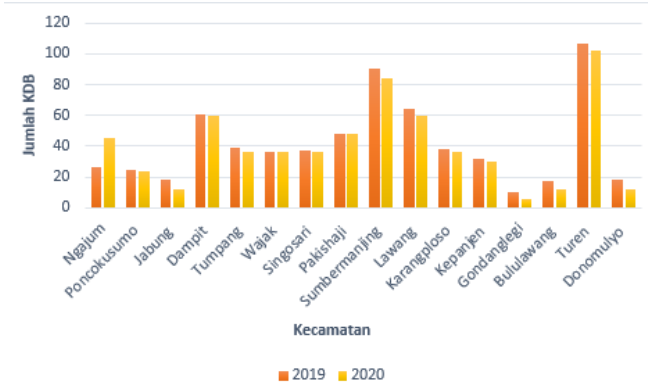
Hasil perbandingan RMSE menunjukkan bahwa pada dataran tinggi metode *Random Forest Regression* memiliki nilai RMSE paling kecil yaitu 1.344. . Pada dataran sedang dan dataran rendah, nilai RMSE *Random Forest Regression* cukup baik dimana nilai RMSE yang dihasilkan tergolong bukan yang terkecil namun juga bukan yang terbesar.

6.7 Analisis Hasil Percobaan

Dalam melakukan analisis hasil percobaan, terdapat 2 jenis analisis yaitu analisis terhadap sudut pandang dari segi manajerial dan analisis dari segi metode yang digunakan pada penelitian tugas akhir ini.

6.6.1. Analisis Segi Manajerial

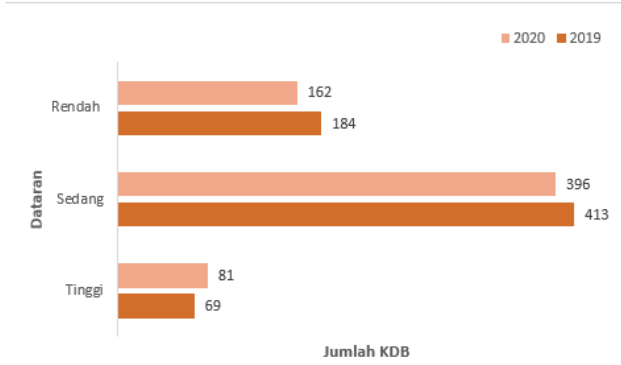
Proses analisis terhadap segi manajerial bertujuan untuk memberikan informasi kepada pihak Dinas Kesehatan Kabupaten Malang terhadap jumlah Kasus Demam Berdarah (KDB) di setiap kecamatan pada periode mendatang. Hasil peramalan jumlah KDB dari 16 kecamatan di Kabupaten Malang dapat dilihat pada Gambar 6.28.



Gambar 6.28. Hasil peramalan jumlah KDB dari 16 kecamatan di Kabupaten Malang

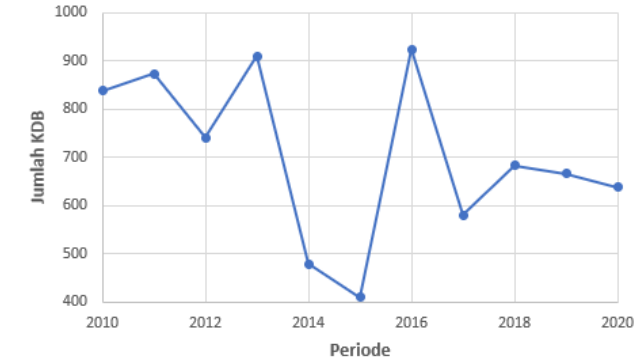
Pada Gambar 6.28 dapat diketahui bahwa jumlah KDB pada tahun 2020 cenderung menurun dibandingkan tahun 2019. Namun, kecamatan Ngajum mengalami peningkatan jumlah KDB dari 26 kasus menjadi 45 kasus pada tahun 2020. Selain itu, jika dilihat dari jumlah KDB paling banyak terdapat pada kecamatan Turen dan Sumbermanjing. Hal ini dapat menjadi perhatian khusus oleh pihak Dinas Kesehatan Malang untuk melakukan tindakan pencegahan agar jumlah KDB menurun, terutama pada kecamatan Turen dan Sumbermanjing serta kecamatan Ngajum yang akan terjadi peningkatan jumlah KDB pada tahun 2020. Jumlah KDB setiap dataran ditunjukkan pada Gambar 6.29.

Dataran tinggi terdiri dari kecamatan Ngajum, Poncokusumo, dan Jabung. Peningkatan jumlah KDB kecamatan Ngajum pada tahun 2020 juga akan memengaruhi jumlah KDB di dataran tinggi. Jumlah KDB di dataran sedang dan rendah mengalami penurunan pada tahun 2020.



Gambar 6.29. Jumlah KDB setiap dataran di Kabupaten Malang

Jumlah kecamatan pada dataran sedang lebih banyak dibandingkan dataran lain sehingga jumlah KDB di dataran sedang lebih tinggi. Dataran sedang terdiri dari kecamatan Dampit, Tumpang, Wajak, Singosari, Pakishaji, Sumbermanjing, Lawang dan Karangploso. Sedangkan dataran rendah terdiri dari kecamatan Kepanjen, Gondanglegi, Bululawang, Turen, dan Donomulyo. Untuk secara keseluruhan jumlah KDB di Kabupaten Malang dari periode 2010 sampai 2020 ditunjukkan pada Gambar 6.30.



Gambar 6.30. Jumlah KDB di Kabupaten Malang dari periode 2010 sampai 2020

Pada Gambar 6.30 menunjukkan bahwa jumlah KDB di Kabupaten Malang diperkirakan dari periode 2019 sampai 2020 mengalami penurunan. Jumlah KDB pada periode 2014 menurun drastis tetapi pada periode 2016 terjadi peningkatan kasus yang tinggi dibandingkan dengan periode sebelumnya dan periode 2018. Penurunan kasus pada periode 2019 sampai 2020 dapat terjadi jika didukung dengan berbagai program pemerintahan dalam melakukan pencegahan jumlah KDB di Kabupaten Malang.

6.6.2. Analisis Segi Metode

Hasil model terbaik yang dihasilkan dari setiap dataran memiliki skenario pembentukan model dan nilai RMSE yang berbeda. Perbedaan nilai RMSE yang dihasilkan dapat dilihat dari perbedaan *range* data yang digunakan saat melakukan pembentukan model. Hasil model terbaik pada kecamatan pembentuk model di setiap dataran ditunjukkan pada Tabel 6.30.

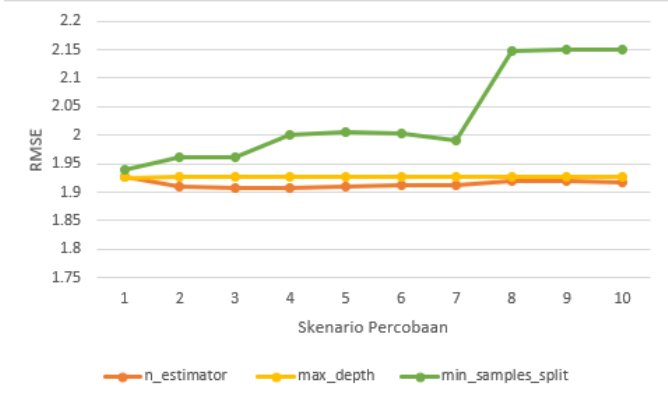
Tabel 6.30. Hasil model terbaik pada kecamatan pembentuk model di setiap dataran

Kecamatan Pembentuk Model	Skenario IV	RMSE Terbaik	SD KDB
Ngajum (Dataran Tinggi)	Top 3 IV Lag 12	1.3437	3.0605
Pakishaji (Dataran Sedang)	No IV Lag 2	2.2568	5.1209
Kepanjen (Dataran Rendah)	No IV Lag 2	2.5694	6.8692

Pada Tabel 6.30 dapat diketahui bahwa nilai RMSE terkecil terdapat pada kecamatan Ngajum. Perbedaan nilai RMSE pada kecamatan Pakishaji dan Kepanjen memiliki perbedaan yang tidak cukup jauh dibandingkan dengan kecamatan Ngajum. Hal ini dikarenakan persebaran data yang berbeda di setiap kecamatan pembentuk model. Persebaran data dapat dilihat dari nilai standar deviasi jumlah KDB setiap kecamatan pembentuk model. Nilai standar deviasi pada kecamatan Ngajum sebesar

3.0605 yang merupakan nilai terkecil dibandingkan dengan 2 kecamatan lainnya.
Perbedaan dari persebaran data akan memengaruhi nilai RMSE yang didapatkan saat proses pembentukan model. Selain itu, nilai parameter yang digunakan pada metode *Random Forest Regression* juga dapat memengaruhi nilai RMSE yang didapatkan. Untuk melihat parameter mana yang lebih sensitif terhadap perubahan nilai, maka pada penelitian tugas akhir ini dilakukan percobaan.

Setiap percobaan dilakukan untuk melihat perubahan nilai RMSE pada 3 parameter yaitu parameter *n_estimator, max_depth, dan min_samples_split.* Nilai parameter yang digunakan untuk melakukan percobaan dapat dilihat pada lampiran F. Sedangkan hasil perubahan nilai RMSE saat menggunakan nilai yang berbeda dari setiap parameter ditunjukkan pada Gambar 6.31.



Gambar 6.31. Hasil perubahan nilai RMSE saat menggunakan nilai yang berbeda dari setiap parameter

Pada Gambar 6.31 terlihat bahwa parameter *min_samples_split* merupakan parameter yang sensitif terhadap perubahan nilai parameter. Dalam melakukan percobaan sebanyak 10 kali, pada penelitian tugas akhir ini menggunakan nilai parameter *min_samples_split* yang berbeda pada setiap percobaan.

Sedangkan parameter *n_estimator* dan *max_depth* hasil nilai RMSE dari 10 percobaan tidak terlalu banyak perubahan.

6.8 Kesimpulan Hasil Percobaan

Hasil peramalan jumlah KDB dengan menggunakan metode *Random Forest Regression* dapat disimpulkan sebagai berikut:

1. Pada dataran tinggi, model terbaik memiliki nilai RMSE terkecil sebesar 1.3437 dan nilai SMAPE terkecil sebesar 24.14% yang terdapat pada pembentukan model dengan 70% data *training* dan 30% data *testing* pada skenario Top 3 IV lag 12. Nilai parameter dari model terbaik yaitu jumlah n_estimator sebanyak 1000, jumlah min_samples_split sebanyak 12, dan jumlah max_depth sebanyak 10. Model ini memperhatikan *Independent variable* curah hujan, jumlah penduduk, suhu udara dan variabel lag.
2. Model terbaik yang dihasilkan pada dataran tinggi merupakan model yang robust terhadap pembagian data namun tidak *robust* terhadap kecamatan lain. Jika dilihat berdasarkan nilai rata-rata RMSE, model *robust* terhadap kecamatan lain terdapat pada skenario All IV lag 12 dengan pembagian 70% data *training* dan 30% data *testing* yang memiliki nilai RMSE sebesar 1.8646. Sedangkan model *robust* terhadap pembagian data, rata-rata nilai RMSE terkecil sebesar 1.6459 terdapat pada skenario Top 3 IV lag 12 dengan pembagian 70% data *training* dan 30% data *testing.* Model *robust* terhadap pembagian data memperhatikan *Independent variable* angka bebas jentik, curah hujan, kelembapan udara dan variabel lag.
3. Pada dataran sedang, model terbaik memiliki nilai RMSE terkecil sebesar 2.5964 dan nilai SMAPE sebesar 35.55% dengan pembagian 80% data training dan 20% data testing. Model terbaik yang dihasilkan di dataran sedang berada pada skenario No IV dengan jumlah lag 2 dengan parameter jumlah n_estimator sebanyak 200, jumlah

min_samples_split sebanyak 16, dan jumlah max_depth sebanyak 10. Model ini hanya melibatkan variabel lag sebagai *input*. Namun, nilai SMAPE terkecil dihasilkan bukan dari model terbaik melainkan dari pembentukan model dengan pembagian 70% data *training* dan 30% data *testing* sebesar 33.69%. Model ini didapatkan dari skenario Top 1 IV lag 3 dengan nilai RMSE sebesar 2.7650 dan memperhatikan *independent variable* angka bebas jentik dan variabel lag.

4. Model terbaik yang dihasilkan pada dataran sedang merupakan model yang *robust* terhadap kecamatan lain namun tidak *robust* terhadap pembagian data. Jika dilihat dari rata-rata nilai RMSE pada semua kecamatan, model *robust* berada pada skenario No IV lag 2 dengan pembagian 80% data *training* dan 20% data *testing* yang memiliki nilai RMSE sebesar 4.7097 . Sedangkan model *robust* terhadap pembagian data berada pada skenario Top 1 IV lag 3 dengan pembagian data sebesar 70% data *training* dan 30% data *testing.* Model ini memperhatikan *Independent variable* variabel angka bebas jentik dan variabel lag.

5. Pada dataran rendah model terbaik memiliki nilai RMSE terkecil sebesar 2.2568 dan nilai SMAPE sebesar 23.52% terdapat pada pembagian 70% data *training* dan 30% data *testing*. Model terbaik yang dihasilkan di dataran rendah berada pada skenario No IV dengan jumlah lag 2 dengan parameter jumlah n_estimator sebanyak 700, jumlah min_samples_split sebanyak 16, dan jumlah max_depth sebanyak 5. Namun, nilai SMAPE terkecil dihasilkan bukan dari model terbaik melainkan terdapat pada pembentukan model 80% data *training* dan 20% data *testing* sebesar 22.52%. Model ini dihasilkan dari skenario Top 1 IV lag 2 dengan nilai RMSE sebesar 2.3046 dan memperhatikan *Independent variable* jumlah penduduk dan variabel lag.

6. Model terbaik yang dihasilkan pada dataran rendah bukan model yang *robust* terhadap kecamatan lain dan pembagian data yang berbeda. Jika dilihat dari rata-rata nilai RMSE terkecil, model *robust* terhadap kecamatan lain dan pembagian data terdapat pada pembagian 80% data *training* dan 20% data *testing* dengan skenario Top 2 IV lag 12 yang memiliki nilai rata-rata RMSE sebesar 3.7590 dan 2.5151. Model ini memperhatikan *Independent* jumlah penduduk, kelembapan udara dan variabel lag.

7. Hasil pengujian model terbaik di setiap dataran tidak signifikan terhadap proporsi pembagian data dan kecamatan lain. Model terbaik yang dipilih berdasarkan nilai RMSE terkecil pada *data testing* belum tentu model yang *robust* terhadap proporsi pembagian data dan kecamatan lain.

8. Pembentukan model dengan menggunakan skenario pembagian data 80%:20% dan 70%:30% memiliki nilai RMSE lebih kecil dibandingkan dengan 60%:40% dan 50%:50%.

9. Setiap model terbaik yang dihasilkan dari proses pembentukan model memiliki jumlah lag dan nilai parameter optimal yang berbeda-beda.

10. Model terbaik yang dihasilkan untuk seluruh kecamatan memiliki nilai median sebesar 2.9675 pada pembentukan model dengan proporsi data 70%:30% dan skenario All IV Lag 12 di kecamatan Ngajum (dataran tinggi). Model ini memperhatikan semua independent variable yaitu suhu, kecepatan angin, kelembapan udara, curah hujan, angka bebas jentik, jumlah penduduk, dan variabel lag 12.

11. Model yang memperhatikan *independent variabel* termasuk variabel lag memiliki tingkat akurasi lebih baik dibandingkan dengan model yang hanya memperhatikan variabel lag tanpa *independent variable*. Hasil model terbaik di dataran tinggi dengan memperhatikan

independent variable memiliki nilai RMSE terkecil dibandingkan dengan dataran sedang dan rendah yang hanya memperhatikan variabel lag saja. Selain itu, model *robust* terhadap kecamatan lain, pembagian data serta model terbaik untuk seluruh kecamatan dihasilkan dari model yang memperhatikan *independent variable* yang mana memiliki tingkat akurasi lebih baik dibandingkan model yang tanpa memperhatikan *independent variable*.

12. Parameter *min_samples_split* pada metode *Random Forest Regression* yang digunakan pada penelitian tugas akhir ini lebih sensitif dibandingkan dengan parameter *n_estimator* dan *max_depth*.

13. Perbedaan dari persebaran dan variasi nilai pada data kecamatan pembentuk model akan memengaruhi nilai RMSE yang dihasilkan.

Halaman ini sengaja dikosongkan

BAB VII
KESIMPULAN DAN SARAN

Pada bab ini dibahas mengenai kesimpulan dari semua proses yang telah dilakukan dan beberapa saran yang dapat diberikan untuk pengembangan yang lebih baik.

7.1. Kesimpulan

Kesimpulan yang didapatkan dari proses pengerjaan tugas akhir yang telah dilakukan antara lain :

1. Pada dataran tinggi model terbaik yang digunakan terdapat pada pembentukan model dengan pembagian 70% data *training* dan 30% data *testing*. Model terbaik yang dihasilkan di dataran tinggi berada pada skenario Top 3 IV lag 12 dengan nilai RMSE terkecil sebesar 1.3437 dengan memperhatikan *independent variable* curah hujan, jumlah penduduk, suhu udara dan variabel lag.

2. Pada dataran sedang model terbaik yang digunakan terdapat pada pembentukan model dengan pembagian 80% data *training* dan 20% data *testing.* Model terbaik yang dihasilkan di dataran sedang berada pada skenario No IV dengan jumlah lag 2 dengan nilai RMSE terkecil sebesar 2.5964. Model ini hanya melibatkan variabel lag sebagai *input.*

3. Pada dataran rendah model terbaik yang digunakan terdapat pada pembentukan model dengan pembagian 70% data *training* dan 30% data *testing.* Model terbaik yang dihasilkan di dataran rendah berada pada skenario No IV dengan jumlah lag 2 dengan nilai RMSE terkecil sebesar 2.2568 dan hanya memperhatikan variabel lag saja.

4. Jumlah KDB pada tahun 2020 cenderung menurun dibandingkan tahun 2019 pada tingkat Kabupaten Malang. Penurunan jumlah KDB dapat membuktikan bahwa program pemerintah untuk mencegah kasus KDB di

Kabupaten Malang berhasil. Meskipun jumlah KDB turun pada tahun 2020, jika dilihat berdasarkan dataran terjadi peningkatan jumlah KDB di dataran tinggi dari 69 kasus pada tahun 2019 menjadi 81 kasus pada tahun 2020. Hal ini disebabkan kecamatan Ngajum mengalami peningkatan jumlah KDB pada tahun 2020 sehingga dapat menjadi perhatian khusus bagi pihak Dinas Kesehatan Kabupaten Malang.

5. Mengapa model terbaik di setiap dataran tidak bisa digeneralisasi untuk semua dataran?

7.2. Saran

Dari pengerjaan tugas akhir ini terdapat beberapa saran untuk pengembangan dalam penelitian selanjutnya.

1. Untuk penelitian selanjutnya dapat melibatkan parameter lain dari metode *Random Forest Regression* seperti min_samples_leaf. Selain itu, perlu dilakukan proses analisis parameter mana saja yang dapat menghasilkan nilai *Root Mean Square Error* (RMSE) terkecil dalam pembentukan model terbaik.
2. Memperbanyak nilai parameter yang digunakan sehingga dapat mengetahui model terbaik dihasilkan dari *range* nilai parameter yang kecil atau besar. Salah satu contoh yaitu memperkecil *range* nilai parameter n_estimator dan memperbesar *range* nilai parameter min_samples_split. Hasil dari percobaan tersebut dapat dilakukan analisis pengaruh *range* nilai dari parameter yang digunakan.
3. Melakukan percobaan dengan melibatkan semua *independent variable* sehingga tidak hanya berdasarkan pada nilai korelasi yang tinggi. Namun, percobaan dilakukan dengan menggunakan satu per satu *independent variable* seperti variabel suhu udara saja dengan jumlah KDB Nilai korelasi yang tinggi tidak menjamin

independent variable tersebut berpengaruh terhadap jumlah KDB. Jika dilakukan percobaan dengan menggunakan setiap *independent variable* maka akan terlihat variabel mana yang paling memengaruhi jumlah KDB.

Halaman ini sengaja dikosongkan

DAFTAR PUSTAKA

[1] N. Amuche, E. Emmanuel, and N. Innocent, "Profile of dengue hepatitis in children from India and its correlation with WHO dengue case classification," *Behaviour*, vol. 9, no. 6, p. 10, 2016.

[2] A. Gama T and F. Betty R, "Analisis Faktor Risiko Kejadian Demam Berdarah Dengue di Desa Mojosongo Kabupaten Boyolali," *Fak. Ilmu Kesehat. Univ. Muhammadyah Surakarta*, vol. 5, pp. 1–9, 2010.

[3] Masrizal and N. P. Sari, "Analisis Kasus DBD Berdasarkan Unsur Iklim dan Kepadatan Penduduk Melalui Pendekatan GIS di Tanah Datar," *J. Kesehat. Masy. Andalas*, vol. 10, no. 2, pp. 166–171, 2016.

[4] "Puncak Musim Hujan, Kabupaten Malang Paling Rawan Terjangkit DBD | MalangTIMES." [Online]. Available: https://www.malangtimes.com/baca/35229/20190120/111000/puncak-musim-hujan-kabupaten-malang-paling-rawan-terjangkit-dbd. [Accessed: 06-Nov-2019].

[5] T. Petukhova, D. Ojkic, B. Mcewen, R. Deardon, and Z. Poljak, "Assessment of autoregressive integrated moving average (ARIMA), generalized linear autoregressive moving average (GLARMA), and random forest (RF) time series regression models for predicting influenza A virus frequency in swine in Ontario , Canada," pp. 1–17, 2018.

[6] M. J. Kane, N. Price, M. Scotch, and P. Rabinowitz, "Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks," 2014.

[7] N. J. Johannesen, M. Kolhe, and M. Goodwin, "Relative evaluation of regression tools for urban area electrical energy demand forecasting," *J. Clean. Prod.*, vol. 218, pp. 555–564, 2019.

[8] T. M. Carvajal, K. M. Viacrusis, L. F. T. Hernandez, H. T. Ho, D. M. Amalin, and K. Watanabe, "Machine learning methods reveal the temporal pattern of dengue

incidence using meteorological factors in metropolitan Manila, Philippines," *BMC Infect. Dis.*, vol. 18, no. 1, pp. 1–15, 2018.

[9] E. Java, "Forecasting the Number of Dengue Fever Cases in Malang Regency Indonesia," vol. 95, no. 1, 2017.

[10] "Kementerian Kesehatan Republik Indonesia." [Online]. Available: https://www.kemkes.go.id/article/view/19011800001/kemenkes-imbau-seluruh-daerah-siaga-dbd.html. [Accessed: 06-Nov-2019].

[11] V. Kotu and B. Deshpande, *Data Science (Concepts and Practice)*, Second Edi. 2019.

[12] I. Pratama, A. E. Permanasari, I. Ardiyanto, and R. Indrayani, "A Review of Missing Values Handling Methods on Time-Series Data," no. April 2018, 2016.

[13] X. Qiu, L. Zhang, P. Nagaratnam, and G. A. J. Amaratunga, "Oblique random forest ensemble via Least Square Estimation for time series forecasting," *Inf. Sci. (Ny).*, vol. 420, pp. 249–262, 2017.

[14] J. Ong *et al.*, "Mapping dengue risk in Singapore using Random Forest," pp. 1–12, 2018.

[15] C. Chen, J. Twycross, and J. M. Garibaldi, "A new accuracy measure based on bounded relative error for time series forecasting," *PLoS One*, vol. 12, no. 3, pp. 1–23, 2017.

[16] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, 2014.

[17] I. Bou-hamad and I. Jamali, "Research in International Business and Finance Forecasting financial time-series using data mining models : A simulation study ⋆," *Res. Int. Bus. Financ.*, vol. 51, no. August 2019, p. 101072, 2020.

[18] M. Lawrence, M. O'Connor, and B. Edmundson, "Field study of sales forecasting accuracy and processes," *Eur.*

J. Oper. Res., vol. 122, no. 1, pp. 151–160, 2000.

[19] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and Machine Learning forecasting methods : Concerns and ways forward," no. Ml, pp. 1–26, 2018.

[20] Shallu and R. Mehra, "Breast cancer histology images classification: Training from scratch or transfer learning?," *ICT Express*, vol. 4, no. 4, pp. 247–254, 2018.

[21] "Random Forest Regression in Python - GeeksforGeeks." [Online]. Available: https://www.geeksforgeeks.org/random-forest-regression-in-python/. [Accessed: 08-Mar-2020].

[22] J. M. Scavuzzo *et al.*, "Modeling Dengue vector population using remotely sensed data and machine learning," *Acta Trop.*, vol. 185, no. October 2017, pp. 167–175, 2018.

[23] X. Zhang, Y. Liu, M. Yang, T. Zhang, A. A. Young, and X. Li, "Comparative Study of Four Time Series Methods in Forecasting Typhoid Fever Incidence in China," *PLoS One*, vol. 8, no. 5, 2013.

[24] "Introduction to Boosted Trees — xgboost 1.2.0-SNAPSHOT documentation." [Online]. Available: https://xgboost.readthedocs.io/en/latest/tutorials/model.html. [Accessed: 24-Jun-2020].

[25] "1.10. Decision Trees — scikit-learn 0.23.1 documentation." [Online]. Available: https://scikit-learn.org/stable/modules/tree.html#regression. [Accessed: 24-Jun-2020].

[26] "Cross-Validation strategies for Time Series forecasting [Tutorial] | Packt Hub." [Online]. Available: https://hub.packtpub.com/cross-validation-strategies-for-time-series-forecasting-tutorial/. [Accessed: 31-May-2020].

[27] "Kelebihan dan Kekurangan Rata-rata, Median dan Modus." [Online]. Available: https://www.rumusstatistik.com/2013/09/kelebihan-dan-kekurangan-rata-rata.html.[Accessed: 12-Jul-2020].

BIODATA PENULIS

Penulis bernama lengkap Firin Handayani dilahirkan di Lumajang, 14 Juni 1997. Penulis telah menempuh pendidikan formal sejak tahun 2004, yaitu SDN Gondoruso 02 pada tingkat sekolah dasar, SMPN 2 Pasirian pada tingkat sekolah menengah pertama, dan SMAN 1 Tempeh pada tingkat sekolah menengah atas.

Pada tahun 2016 pasca kelulusan SMA, penulis melanjutkan pendidikan di Jurusan Sistem Informasi Fakultas Teknologi Informasi dan Komunikasi – Institut Teknologi Sepuluh Nopember (ITS) Surabaya melalui jalur SNMPTN dan terdaftar sebagai mahasiswa dengan NRP 05211610000006. Di awal perkuliahan penulis telah menjadi penerima beasiswa bidikmisi. Selama menjadi mahasiswa, penulis aktif mengikuti berbagai kegiatan kemahasiswaan. Pada tahun pertama perkuliahan penulis menjadi staff sie Event FTif Festival 2017. Pada tahun kedua menjadi staff Departemen Sosial Masyarakat bagian DiIVsi Social Edu HMSI periode 2017/2018 dan staff sie Event Information System Expo 2017. Pada tahun ketiga menjabat sebagai Kepala DiIVsi Pengajaran dan Pengabdian di Departemen Sosial Masyarakat, panitia staff sie penyisihan pada GemasTIK ke-11 2018, menjadi pemandu integralistik di Gerigi ITS 2018, menjadi Koordinator sie Event BIONIX Information System Expo 2018, dan menjadi staff sie Konsumsi di Bursa Karir ITS ke-37 2019.

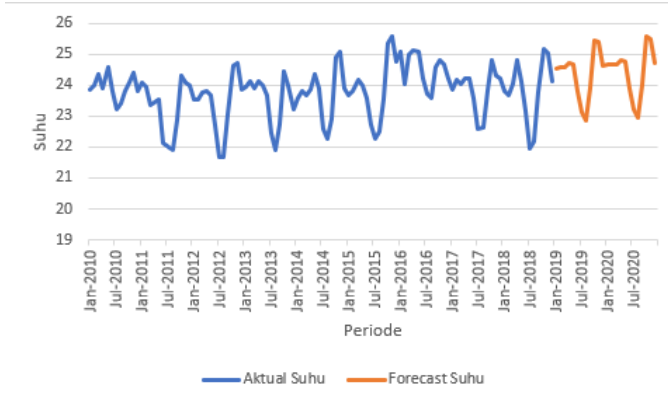
Pada tahun keempat, penulis memiliki ketertarikan di bidang pengolahan data sehingga penulis mengambil topik tugas akhir peramalan data. Topik ini termasuk salah satu bidang minat di Laboratorium Rekayasa Data dan Intelegensi Bisnis (RDIB). Jika ingin memberikan masukan dan saran terkait dengan tugas akhir ini, dapat menghubungi penulis melalui email firinhandayani@gmail.com.

Halaman ini sengaja dikosongkan

LAMPIRAN A

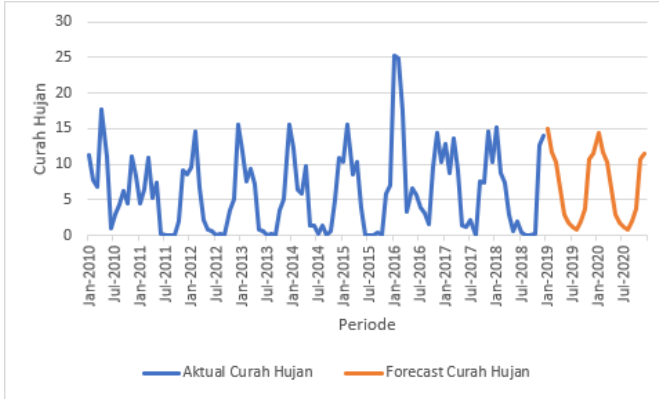
Gambar Grafik Hasil Peramalan Independent Variable

a. Suhu udara



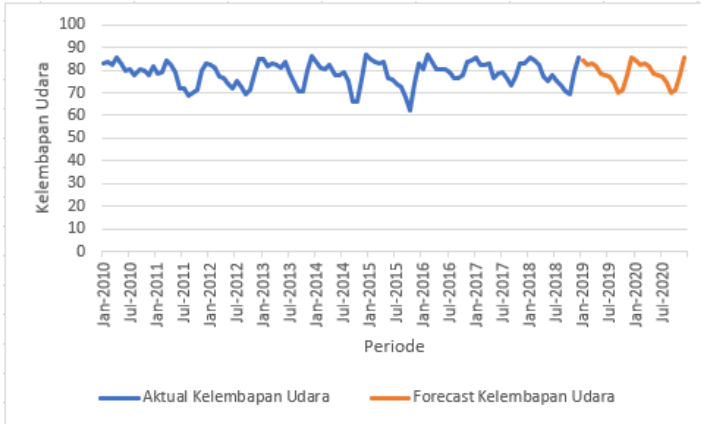
Gambar A.1. Hasil Peramalan Variabel Suhu udara pada Stasiun Karangploso

b. Curah Hujan



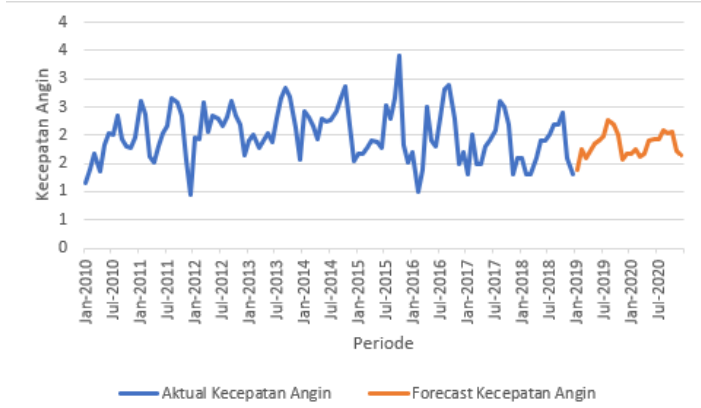
Gambar A.2. Hasil Peramalan Variabel Curah Hujan pada Stasiun Karangploso

c. Kelembapan Udara



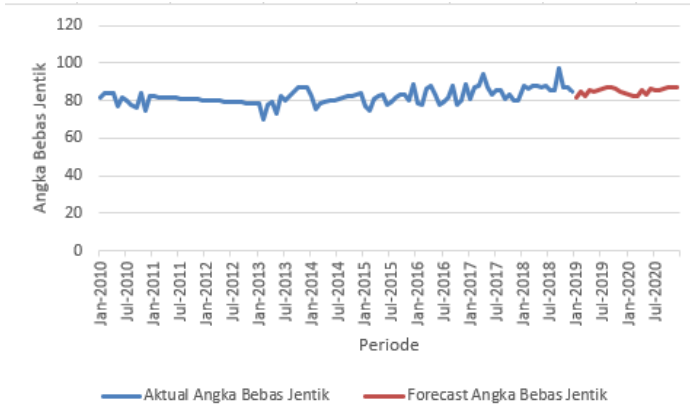
Gambar A.3. Hasil Peramalan Variabel Kelembapan Udara pada Stasiun Karangploso

d. Kecepatan Angin



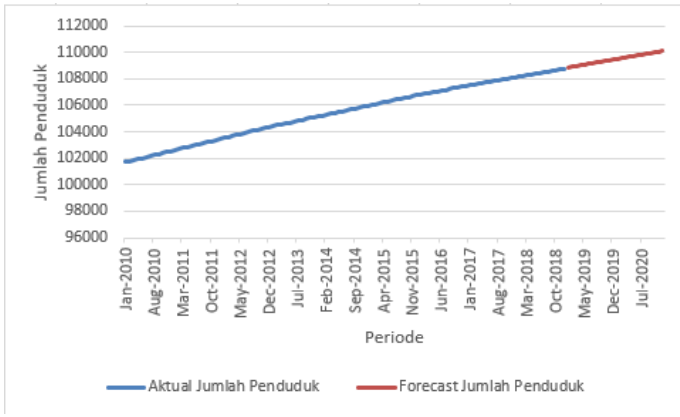
Gambar A.4. Hasil Peramalan Variabel Kecepatan Angin pada Stasiun Karangploso

e. Angka Bebas Jentik



Gambar A.5. Hasil Peramalan Variabel Angka Bebas Jentik pada Kecamatan Kepanjen

f. Jumlah Penduduk



Gambar A.6. Hasil Peramalan Variabel Jumlah Penduduk pada Kecamatan Kepanjen

LAMPIRAN B

Hasil Uji Korelasi Pemilihan Kecamatan Pembentuk Model

2. Dataran Tinggi

Tabel B.1. Hasil Uji Korelasi Pemilihan Kecamatan Pembntuk Model pada Dataran Tinggi

Kecamatan	Poncokusumo	Jabung	Ngajum
Poncokusumo	1	0.2834	0.6879
Jabung	0.2834	1	0.4794
Ngajum	0.6879	0.4794	1
avg	0.4856	0.3814	0.5836

3. Dataran Sedang

Tabel B.2. Hasil Uji Korelasi Pemilihan Kecamatan Pembentuk Model pada Dataran Sedang

Kecamatan	Dampit	Tumpang	Wajak	Singosari	Pakishaji	Sumber manjing	Lawang	Karangploso
Dampit	1	-0.0152	0.4519	0.4762	0.5174	0.6859	0.1888	0.0885
Tumpang	-0.0152	1	0.1122	-0.2217	0.0325	-0.1263	0.2871	0.1296
Wajak	0.4519	0.1122	1	0.1976	0.4879	0.4414	0.5298	0.2344
Singosari	0.4762	-0.2217	0.1976	1	0.3503	0.5502	-0.0003	0.2008
Pakishaji	0.5174	0.0325	0.4879	0.3503	1	0.4252	0.3842	0.4014
Sumbermanjing	0.6859	-0.1263	0.4414	0.5502	0.4252	1	0.2915	0.1350
Lawang	0.1888	0.2871	0.5298	-0.0003	0.3842	0.2915	1	0.4143
Karangploso	0.0885	0.1296	0.2344	0.2008	0.4014	0.1350	0.4143	1
avg	0.3419	0.0283	0.3507	0.2219	0.3713	0.3433	0.2993	0.2291

4. Dataran Rendah

Tabel B.3.Hasil Uji Korelasi Pemilihan Kecamatan Pembentuk Model pada Dataran Rendah

Kecamatan	Kepanjen	GondangLegi	Bululawang	Turen	Donomulyo
Kepanjen	1	0.6438	0.5645	0.8050	0.5172
GondangLegi	0.6438	1	0.3929	0.5433	0.4207
Bululawang	0.5645	0.3929	1	0.3370	0.3977
Turen	0.8050	0.5433	0.3370	1	0.5011
Donomulyo	0.5172	0.4207	0.3977	0.5011	1
avg	0.6326	0.5002	0.4230	0.5466	0.4592

LAMPIRAN C

Hasil Pembentukan Model Semua Skenario

1. Dataran Tinggi

Tabel C.1. Hasil Pembentukan Model Semua Skenario pada Dataran Tinggi

Skenario Pembagian Data	Skenario IV	Lag	n_estimator	Min_Samples_Split	Max_Depth	RMSE Training	RMSE Testing	SMAPE Testing
80-20	All IV	12	300	8	10	1.3965	1.4921	26.3353
	Top 3 IV	12	300	4	10	1.3285	1.4923	26.5342
	Top 2 IV	0	700	12	10	1.5224	1.5054	27.4611
	Top 1 IV	12	600	4	10	1.3017	1.6004	28.5676
	No IV	12	1000	8	10	2.9311	1.7793	26.6702
70-30	All IV	12	1000	12	10	1.4662	1.3698	24.4757
	Top 3 IV	12	1000	12	10	1.4700	1.3437	24.1388
	Top 2 IV	12	1000	12	10	1.4851	1.3573	24.5703
	Top 1 IV	12	100	4	10	1.2712	1.4756	25.3369

	No IV	2	200	8	10	3.3816	1.9101	32.8177
60-40	All IV	6	400	16	10	1.0212	4.3360	36.0567
	Top 3 IV	7	500	16	5	1.0503	4.3269	36.1601
	Top 2 IV	6	200	16	5	1.1208	4.3189	36.8906
	Top 1 IV	2	200	12	5	1.0235	4.5813	42.4339
	No IV	10	500	4	5	1.0265	4.5471	36.0223
50-50	All IV	4	400	12	5	0.9753	3.9909	48.3334
	Top 3 IV	4	400	12	5	0.9954	3.9753	48.5211
	Top 2 IV	4	400	12	5	1.0180	3.9955	48.7139
	Top 1 IV	2	600	8	10	1.0152	4.1187	54.1130
	No IV	10	400	4	5	0.8963	4.1086	46.2696

2. Dataran Sedang

Tabel C.2. Hasil Pembentukan Model Semua Skenario pada Dataran Sedang

Skenario Pembagian Data	Skenario IV	Lag	n_estimator	Min_Samples _Split	Max_Depth	RMSE Training	RMSE Testing	SMAPE Testing
80-20	All IV	2	200	16	5	3.4040	3.6191	39.8748
	Top 3 IV	3	100	8	10	3.0782	3.5460	37.6071
	Top 2 IV	2	200	16	10	3.5644	3.2169	39.3294
	Top 1 IV	2	200	16	10	3.6240	3.0455	37.1631
	No IV	2	200	16	10	4.7973	2.5965	35.5511
70-30	All IV	3	600	16	5	3.6419	3.3201	38.1054
	Top 3 IV	3	100	16	10	3.6296	3.2762	37.9005
	Top 2 IV	3	100	16	10	3.7704	3.3852	39.2626
	Top 1 IV	3	100	4	5	2.9502	2.7650	33.6864
	No IV	1	1000	8	5	5.4102	2.7967	35.4416
60-40	All IV	3	100	16	5	3.6192	4.8775	38.7768
	Top 3 IV	3	100	16	5	3.7950	4.7961	38.1941
	Top 2 IV	2	200	12	5	3.7073	4.5968	38.1174

	Top 1 IV	2	100	12	5	3.6951	4.4976	34.8707
	No IV	2	600	8	5	3.8264	4.9875	33.2589
50-50	All IV	12	400	4	10	2.0597	5.4268	42.3094
	Top 3 IV	3	100	12	5	3.6762	4.8533	36.5648
	Top 2 IV	3	400	12	5	3.9571	4.4053	37.5072
	Top 1 IV	2	100	12	5	4.0120	4.4822	36.7042
	No IV	3	100	8	5	3.8745	4.9543	33.4423

3. Dataran Rendah

Tabel C.3. Hasil Pembentukan Model Semua Skenario pada Dataran Rendah

Skenario Pembagian Data	Skenario IV	Lag	n_estimator	Min_Samples _Split	Max_Depth	RMSE Training	RMSE Testing	SMAPE Testing
80-20	All IV	12	300	16	10	3.8952	2.5978	23.4569
	Top 3 IV	3	200	16	5	4.4606	2.6530	25.5392
	Top 2 IV	12	100	4	5	2.1884	2.6049	23.9412
	Top 1 IV	2	200	16	10	5.0054	2.3047	22.5193
	No IV	2	700	16	5	5.3891	2.5036	24.5240

70-30	All IV	7	100	12	10	3.8147	2.8169	28.2593
	Top 3 IV	7	100	12	10	4.1122	2.6147	26.8789
	Top 2 IV	7	100	12	10	4.5044	3.1431	30.6985
	Top 1 IV	9	500	12	10	4.3313	3.1440	31.6063
	No IV	2	700	16	5	5.7031	2.2569	23.5219
60-40	All IV	12	300	4	5	2.1510	4.3797	29.0793
	Top 3 IV	12	800	4	10	2.2522	4.2714	28.0381
	Top 2 IV	12	600	4	10	2.2103	4.3628	28.1973
	Top 1 IV	0	900	4	15	3.2655	3.9256	30.2993
	No IV	1	500	8	5	5.5423	4.3087	31.8286
50-50	All IV	12	300	4	5	2.2474	4.3642	38.9260
	Top 3 IV	12	700	4	10	2.3769	4.2579	37.4921
	Top 2 IV	12	300	4	10	2.2079	4.4212	40.2199
	Top 1 IV	1	1000	16	5	5.4423	4.1594	30.2530
	No IV	1	400	8	5	5.9891	3.9496	32.9232

LAMPIRAN D

1. Validasi terhadap kecamatan lain

a. Dataran Tinggi

Tabel D.1. Hasil Validasi terhadap Kecamatan Lain pada Dataran Tinggi

Skenario Pembagian Data	Skenario IV	Lag	Nilai RMSE			avg RMSE
			Ngajum	Poncokusumo	Jabung	
80% : 20%	All IV	12	1.4921	2.2796	2.0970	1.9562
	Top 3 IV	12	1.4923	2.2870	2.1482	1.9759
	Top 2 IV	0	1.5054	2.3975	2.4802	2.1277
	Top 1 IV	12	1.6004	2.3205	2.1595	2.0268
	No IV	12	1.7793	2.7351	2.1920	2.2355
70% : 30%	All IV	12	1.3698	2.1991	2.0250	1.8646
	Top 3 IV	12	1.3437	2.2031	2.0711	1.8726
	Top 2 IV	12	1.3573	2.2046	2.0632	1.8750
	Top 1 IV	12	1.4756	2.1793	2.0019	1.8856

	No IV	2	1.9101	2.7381	2.5051	2.3844
60% : 40%	All IV	6	4.3360	2.5843	2.3595	3.0933
	Top 3 IV	7	4.3269	2.6170	2.3912	3.1117
	Top 2 IV	6	4.3189	2.5605	2.3454	3.0749
	Top 1 IV	2	4.5813	2.7651	2.5007	3.2823
	No IV	10	4.5471	2.6307	2.3024	3.1601
50% : 50%	All IV	4	3.9909	2.5906	2.3167	2.9661
	Top 3 IV	4	3.9753	2.5847	2.3147	2.9582
	Top 2 IV	4	3.9955	2.5928	2.3214	2.9699
	Top 1 IV	2	4.1187	2.7484	2.4796	3.1155
	No IV	10	4.1086	2.6259	2.3073	3.0139

b. Dataran Sedang

Tabel D.2. Hasil Validasi terhadap Kecamatan Lain pada Dataran Sedang

Skenario Pembagian Data	Skenario IV	Lag	Nilai RMSE							avg RMSE
			Pakishaji	Dampit	Tumpang	Wajak	Singo sari	Sumber manjing	Lawang	
80% : 20%	All IV	2	3.6191	3.9551	4.7505	3.1531	2.7623	9.1685	7.5066	4.7896
	Top 3 IV	3	3.5460	4.6687	4.9732	3.4617	3.1807	11.2603	7.8727	5.3220
	Top 2 IV	2	3.2169	4.4617	4.5675	3.2403	2.9557	9.5318	7.2977	4.8116
	Top 1 IV	2	3.0455	5.3930	4.5228	3.3720	3.2636	11.6140	8.0041	5.3339
	No IV	2	2.5965	4.5618	4.7728	3.4700	3.1360	8.9158	7.0548	4.7097
70% : 30%	All IV	3	3.3201	4.0800	4.7720	3.2009	2.8448	9.4277	7.5745	4.8360
	Top 3 IV	3	3.2762	4.3772	4.8128	3.3534	3.0556	10.6363	7.7685	5.1043
	Top 2 IV	3	3.3852	4.4800	4.6452	3.2838	3.1473	9.9251	7.4956	4.9735
	Top 1 IV	3	2.7650	5.6678	4.7015	3.5523	3.5684	12.2889	8.1458	5.5496
	No IV	1	2.7967	4.9972	4.9167	3.4175	3.0409	9.5825	7.0187	4.8596
60% : 40%	All IV	3	4.8775	4.5951	5.0097	3.2836	3.0033	9.8876	7.7049	5.2223
	Top 3 IV	3	4.7961	4.8671	4.9886	3.3765	3.1926	10.8695	7.7472	5.4080
	Top 2 IV	2	4.5968	5.0387	4.8513	3.2998	3.2476	9.7889	7.3502	5.1755

	Top 1 IV	2	4.4976	5.4248	4.9755	3.4669	3.1071	10.1441	7.2290	5.2560
	No IV	2	4.9875	4.9727	4.9055	3.5430	3.3558	9.1137	7.1033	5.1657
50% : 50%	All IV	12	5.4268	4.5607	4.7496	3.2168	3.0529	9.9555	8.8175	5.4447
	Top 3 IV	3	4.8533	5.2401	5.2768	3.6390	3.3798	12.1508	8.3604	5.8372
	Top 2 IV	3	4.4053	5.1513	4.7388	3.3802	3.3371	10.2496	7.3338	5.2430
	Top 1 IV	2	4.4822	5.4219	4.8889	3.4800	3.1749	10.1676	7.1121	5.2483
	No IV	3	4.9543	5.2555	4.8370	3.5550	3.3356	9.6348	7.0801	5.2589

c. Dataran Rendah

Tabel D.3. Hasil Validasi terhadap Kecamatan Lain pada Dataran Rendah

Skenario Pembagian Data	Skenario IV	Lag	Nilai RMSE					avg RMSE
			Kepanjen	Turen	Gondanglegi	Donomulyo	Bululawang	
80% : 20%	All IV	12	2.5978	9.5100	2.1542	2.2788	2.4596	3.8001
	Top 3 IV	3	2.6530	13.5675	2.2085	2.2603	2.7288	4.6836
	Top 2 IV	12	2.6049	9.1617	2.1759	2.2725	2.5799	3.7590
	Top 1 IV	2	2.3047	14.8527	2.3080	2.4994	2.9469	4.9823
	No IV	2	2.5036	14.8485	2.3187	2.5488	3.0602	5.0560

70% : 30%	All IV	7	2.8169	10.3935	2.1619	2.1887	2.7087	4.0539
	Top 3 IV	7	2.6147	13.6736	2.3120	2.3140	2.7919	4.7412
	Top 2 IV	7	3.1431	16.9098	2.2770	2.3081	2.9062	5.5088
	Top 1 IV	9	3.1440	17.4431	2.2612	2.3633	3.0159	5.6455
	No IV	2	2.2569	14.9031	2.3164	2.5560	3.0545	5.0174
60% : 40%	All IV	12	4.3797	8.8471	2.0235	2.2566	2.4473	3.9908
	Top 3 IV	12	4.2714	8.9936	2.0223	2.2338	2.5032	4.0049
	Top 2 IV	12	4.3628	8.8992	2.0164	2.2334	2.5366	4.0097
	Top 1 IV	0	3.9256	17.1479	0.2363	2.4205	2.8965	5.3254
	No IV	1	4.3087	16.6975	2.3277	2.6400	3.3465	5.8641
50% : 50%	All IV	12	4.3642	9.0642	2.0468	2.3016	2.5303	4.0614
	Top 3 IV	12	4.2579	9.1270	2.0505	2.3045	2.5458	4.0571
	Top 2 IV	12	4.4212	9.0625	2.0391	2.3094	2.5905	4.0845
	Top 1 IV	1	4.1594	13.9504	2.3226	2.5820	3.2171	5.2463
	No IV	1	3.9496	16.8896	2.3234	2.6146	3.3031	5.8160

2. Validasi terhadap pembagian data

a. Dataran Tinggi

Tabel D.4. Hasil Validasi terhadap Pembagian Data pada Dataran Tinggi

Skenario Pembagian Data	Skenario IV	Lag	RMSE Pembagian Data				avg RMSE	avg SD
			80:20	70:30	60:40	50:50		
80% : 20%	All IV	12	1.4921	1.3333	2.0709	1.8809	1.6943	0.2949
	Top 3 IV	12	1.4923	1.2925	2.0373	1.8442	1.6666	0.2913
	Top 2 IV	0	1.5054	1.5140	2.1296	1.9759	1.7812	0.2769
	Top 1 IV	12	1.6004	1.3778	2.0141	1.8249	1.7043	0.2387
	No IV	12	1.7793	1.7304	3.9006	3.6400	2.7626	1.0121
70% : 30%	All IV	12	1.4419	1.3698	2.0235	1.8438	1.6697	0.2726
	Top 3 IV	12	1.4043	1.3437	2.0033	1.8322	1.6459	0.2793
	Top 2 IV	12	1.4190	1.3573	2.0070	1.8396	1.6557	0.2749
	Top 1 IV	12	1.5344	1.4756	1.9655	1.7833	1.6897	0.1967
	No IV	2	1.9598	1.9101	4.4865	4.0631	3.1049	1.1796
60% : 40%	All IV	6	1.4662	1.4882	4.3360	3.8951	2.7964	1.3284
	Top 3 IV	7	1.4637	1.4813	4.3269	3.8865	2.7896	1.3263
	Top 2 IV	6	1.5175	1.5666	4.3189	3.8832	2.8216	1.2889
	Top 1 IV	2	1.9478	1.8740	4.5813	4.1292	3.1331	1.2329

	No IV	10	1.6945	1.5749	4.5471	4.1581	2.9937	1.3665
50% : 50%	All IV	4	1.3710	1.4887	4.3149	3.9909	2.7914	1.3670
	Top 3 IV	4	1.3580	1.4915	4.2893	3.9753	2.7785	1.3591
	Top 2 IV	4	1.3828	1.5157	4.2600	3.9955	2.7885	1.3433
	Top 1 IV	2	2.0117	1.9158	4.5672	4.1187	3.1534	1.2006
	No IV	10	1.6848	1.5661	4.4807	4.1086	2.9600	1.3417

c. Dataran Sedang

Tabel D.5. Hasil Validasi terhadap Pembagian Data pada Dataran Sedang

Skenario Pembagian Data	Skenario IV	Lag	RMSE Pembagian Data				avg RMSE	avg SD
			80:20	70:30	60:40	50:50		
80% : 20%	All IV	2	3.6191	3.1165	3.2387	3.1066	3.2702	0.2080
	Top 3 IV	3	3.5460	2.9430	3.0401	2.8593	3.0971	0.2670
	Top 2 IV	2	3.2169	2.8415	2.9418	2.8936	2.9735	0.1450
	Top 1 IV	2	3.0455	2.7230	2.8338	2.8339	2.8591	0.1167
	No IV	2	2.5965	2.6142	4.6953	4.4769	3.5957	0.9934
70% : 30%	All IV	3	3.6516	3.3201	3.4918	3.3073	3.4427	0.1409
	Top 3 IV	3	3.6362	3.2762	3.3283	3.1521	3.3482	0.1782
	Top 2 IV	3	3.5606	3.3852	3.3877	3.1837	3.3793	0.1334
	Top 1 IV	3	2.8508	2.7650	2.7995	2.6939	2.7773	0.0570
	No IV	1	2.8350	2.7967	4.8596	4.6777	3.7923	0.9786
60% : 40%	All IV	3	3.7817	3.4101	4.8775	4.4917	4.1402	0.5764
	Top 3 IV	3	3.7752	3.3927	4.7961	4.4221	4.0965	0.5464
	Top 2 IV	2	3.5408	3.5932	4.5968	4.2774	4.0020	0.4498

	Top 1 IV	2	3.5766	3.3165	4.4976	4.2114	3.9005	0.4741
	No IV	2	2.1650	2.7296	4.9875	4.7247	3.6517	1.2244
50% : 50%	All IV	12	4.4008	4.3504	5.3394	5.4268	4.8793	0.5050
	Top 3 IV	3	3.9247	3.5085	4.9777	4.8533	4.3160	0.6188
	Top 2 IV	3	3.6517	3.6193	4.6683	4.4053	4.0861	0.4603
	Top 1 IV	2	3.3000	3.1942	4.5851	4.4822	3.8904	0.6454
	No IV	3	2.3597	2.6653	5.0163	4.9543	3.7489	1.2413

d. Dataran Rendah

Tabel D.6. Hasil Validasi terhadap Pembagian Data pada Dataran Rendah

Skenario Pembagian Data	Skenario IV	Lag	RMSE Pembagian Data				avg RMSE	avg SD
			80:20	70:30	60:40	50:50		
80% : 20%	All IV	12	2.5978	2.8647	3.2840	3.0466	2.9483	0.2511
	Top 3 IV	3	2.6530	2.3433	2.7300	2.8670	2.6483	0.1921
	Top 2 IV	12	2.6049	2.5168	2.5656	2.3731	2.5151	0.0877
	Top 1 IV	2	2.3047	2.1849	3.9316	3.6196	3.0102	0.7745
	No IV	2	2.5036	2.1955	4.0980	3.8750	3.1680	0.8295
70% : 30%	All IV	7	3.0255	2.8169	3.1426	3.0745	3.0149	0.1216
	Top 3 IV	7	2.8781	2.6147	2.9329	2.9362	2.8405	0.1324
	Top 2 IV	7	2.9264	3.1431	4.0470	3.8248	3.4853	0.4638
	Top 1 IV	9	2.9228	3.1440	4.1289	3.8136	3.5023	0.4883
	No IV	2	2.5504	2.2569	4.0847	3.8654	3.1894	0.7963
60% : 40%	All IV	12	3.1039	3.1463	4.3797	3.9946	3.6561	0.5484
	Top 3 IV	12	2.9808	2.9004	4.2714	3.8894	3.5105	0.5864
	Top 2 IV	12	2.9655	2.8724	4.3628	3.9716	3.5431	0.6401

	Top 1 IV	0	2.5103	2.3741	3.9256	3.6146	3.1061	0.6747
	No IV	1	2.6655	2.4911	4.3087	3.9868	3.3630	0.7953
50% : 50%	All IV	12	3.4774	3.5144	4.6508	4.3642	4.0017	0.5160
	Top 3 IV	12	3.3946	3.3324	4.5250	4.2579	3.8775	0.5231
	Top 2 IV	12	3.3969	3.3239	4.6996	4.4212	3.9604	0.6086
	Top 1 IV	1	2.9378	2.4887	4.5099	4.1594	3.5239	0.8354
	No IV	1	2.3644	2.2589	4.2218	3.9496	3.1987	0.8930

LAMPIRAN E

Hasil Peramalan 24 Periode Mendatang

1. Dataran Tinggi

Tabel E.1. Hasil Peramalan pada Dataran Tinggi

a. Kecamatan Poncokusumo

Periode	Jumlah KDB	Periode	Jumlah KDB
Jan-2019	2	Jan-2020	2
Feb-2019	2	Feb-2020	2
Mar-2019	2	Mar-2020	2
Apr-2019	2	Apr-2020	2
Mei-2019	3	Mei-2020	2
Jun-2019	2	Jun-2020	2
Jul-2019	2	Jul-2020	2
Agt-2019	2	Agt-2020	2
Sep-2019	2	Sep-2020	2
Okt-2019	2	Okt-2020	2
Nop-2019	2	Nop-2020	2
Des-2019	2	Des-2020	2

b. Kecamatan Jabung

Periode	Jumlah KDB
Jan-2019	2
Feb-2019	1
Mar-2019	2
Apr-2019	2
Mei-2019	2
Jun-2019	2
Jul-2019	1
Agt-2019	1
Sep-2019	1
Okt-2019	2
Nop-2019	1
Des-2019	1

Periode	Jumlah KDB
Jan-2020	1
Feb-2020	1
Mar-2020	1
Apr-2020	1
Mei-2020	1
Jun-2020	1
Jul-2020	1
Agt-2020	1
Sep-2020	1
Okt-2020	1
Nop-2020	1
Des-2020	1

2. Dataran Sedang

Tabel E.2. Hasil Peramalan pada Dataran Sedang

a. Kecamatan Tumpang

Periode	Jumlah KDB	Periode	Jumlah KDB
Jan-2019	3	Jan-2020	3
Feb-2019	5	Feb-2020	3
Mar-2019	3	Mar-2020	3
Apr-2019	4	Apr-2020	3
Mei-2019	3	Mei-2020	3
Jun-2019	3	Jun-2020	3
Jul-2019	3	Jul-2020	3
Agt-2019	3	Agt-2020	3
Sep-2019	3	Sep-2020	3
Okt-2019	3	Okt-2020	3
Nop-2019	3	Nop-2020	3
Des-2019	3	Des-2020	3

b. Kecamatan Wajak

Periode	Jumlah KDB
Jan-2019	3
Feb-2019	3
Mar-2019	3
Apr-2019	3
Mei-2019	3
Jun-2019	3
Jul-2019	3
Agt-2019	3
Sep-2019	3
Okt-2019	3
Nop-2019	3
Des-2019	3

Periode	Jumlah KDB
Jan-2020	3
Feb-2020	3
Mar-2020	3
Apr-2020	3
Mei-2020	3
Jun-2020	3
Jul-2020	3
Agt-2020	3
Sep-2020	3
Okt-2020	3
Nop-2020	3
Des-2020	3

c. Kecamatan Singosari

Periode	Jumlah KDB
Jan-2019	4
Feb-2019	3
Mar-2019	3
Apr-2019	3
Mei-2019	3
Jun-2019	3
Jul-2019	3
Agt-2019	3
Sep-2019	3
Okt-2019	3
Nop-2019	3

Periode	Jumlah KDB
Jan-2020	3
Feb-2020	3
Mar-2020	3
Apr-2020	3
Mei-2020	3
Jun-2020	3
Jul-2020	3
Agt-2020	3
Sep-2020	3
Okt-2020	3
Nop-2020	3

Des-2019	3	Des-2020	3

d. Kecamatan Dampit

Periode	Jumlah KDB	Periode	Jumlah KDB
Jan-2019	6	Jan-2020	5
Feb-2019	5	Feb-2020	5
Mar-2019	5	Mar-2020	5
Apr-2019	5	Apr-2020	5
Mei-2019	5	Mei-2020	5
Jun-2019	5	Jun-2020	5
Jul-2019	5	Jul-2020	5
Agt-2019	5	Agt-2020	5
Sep-2019	5	Sep-2020	5
Okt-2019	5	Okt-2020	5
Nop-2019	5	Nop-2020	5
Des-2019	5	Des-2020	5

e. Kecamatan Sumbermanjing

Periode	Jumlah KDB	Periode	Jumlah KDB
Jan-2019	6	Jan-2020	7
Feb-2019	14	Feb-2020	7
Mar-2019	7	Mar-2020	7
Apr-2019	7	Apr-2020	7
Mei-2019	7	Mei-2020	7
Jun-2019	7	Jun-2020	7
Jul-2019	7	Jul-2020	7
Agt-2019	7	Agt-2020	7
Sep-2019	7	Sep-2020	7

Okt-2019	7
Nop-2019	7
Des-2019	7

Okt-2020	7
Nop-2020	7
Des-2020	7

f. Kecamatan Lawang

Periode	Jumlah KDB
Jan-2019	4
Feb-2019	8
Mar-2019	6
Apr-2019	6
Mei-2019	5
Jun-2019	5
Jul-2019	5
Agt-2019	5
Sep-2019	5
Okt-2019	5
Nop-2019	5
Des-2019	5

Periode	Jumlah KDB
Jan-2020	5
Feb-2020	5
Mar-2020	5
Apr-2020	5
Mei-2020	5
Jun-2020	5
Jul-2020	5
Agt-2020	5
Sep-2020	5
Okt-2020	5
Nop-2020	5
Des-2020	5

g. Kecamatan Karangploso

Periode	Jumlah KDB
Jan-2019	4
Feb-2019	3
Mar-2019	4
Apr-2019	3
Mei-2019	3
Jun-2019	3
Jul-2019	3
Agt-2019	3

Periode	Jumlah KDB
Jan-2020	3
Feb-2020	3
Mar-2020	3
Apr-2020	3
Mei-2020	3
Jun-2020	3
Jul-2020	3
Agt-2020	3

Sep-2019	3	Sep-2020	3
Okt-2019	3	Okt-2020	3
Nop-2019	3	Nop-2020	3
Des-2019	3	Des-2020	3

3. Dataran Rendah

Tabel E.3. Hasil Peramalan pada Dataran Rendah

a. Kecamatan Gondanglegi

Periode	Jumlah KDB	Periode	Jumlah KDB
Jan-2019	2	Jan-2020	0
Feb-2019	1	Feb-2020	1
Mar-2019	1	Mar-2020	1
Apr-2019	1	Apr-2020	0
Mei-2019	1	Mei-2020	0
Jun-2019	1	Jun-2020	1
Jul-2019	1	Jul-2020	1
Agt-2019	0	Agt-2020	0
Sep-2019	0	Sep-2020	0
Okt-2019	1	Okt-2020	1
Nop-2019	1	Nop-2020	1
Des-2019	0	Des-2020	0

b. Kecamatan Bululawang

Periode	Jumlah KDB
Jan-2019	3
Feb-2019	3
Mar-2019	2
Apr-2019	1
Mei-2019	1
Jun-2019	1
Jul-2019	1
Agt-2019	1
Sep-2019	1
Okt-2019	1
Nop-2019	1
Des-2019	1

Periode	Jumlah KDB
Jan-2020	1
Feb-2020	1
Mar-2020	1
Apr-2020	1
Mei-2020	1
Jun-2020	1
Jul-2020	1
Agt-2020	1
Sep-2020	1
Okt-2020	1
Nop-2020	1
Des-2020	1

c. Kecamatan Turen

Periode	Jumlah KDB
Jan-2019	15
Feb-2019	6
Mar-2019	7
Apr-2019	11
Mei-2019	11
Jun-2019	6
Jul-2019	6
Agt-2019	11
Sep-2019	11
Okt-2019	6
Nop-2019	6
Des-2019	11

Periode	Jumlah KDB
Jan-2020	11
Feb-2020	6
Mar-2020	6
Apr-2020	11
Mei-2020	11
Jun-2020	6
Jul-2020	6
Agt-2020	11
Sep-2020	11
Okt-2020	6
Nop-2020	6
Des-2020	11

d. Kecamatan Donomulyo

Periode	Jumlah KDB
Jan-2019	4
Feb-2019	3
Mar-2019	2
Apr-2019	1
Mei-2019	1
Jun-2019	1
Jul-2019	1
Agt-2019	1
Sep-2019	1
Okt-2019	1
Nop-2019	1
Des-2019	1

Periode	Jumlah KDB
Jan-2020	1
Feb-2020	1
Mar-2020	1
Apr-2020	1
Mei-2020	1
Jun-2020	1
Jul-2020	1
Agt-2020	1
Sep-2020	1
Okt-2020	1
Nop-2020	1
Des-2020	1

LAMPIRAN F

Nilai parameter yang digunakan untuk melakukan percobaan dengan tujuan melihat parameter mana yang lebih sensitif terhadap perubahan nilai.

Tabel F.1. Nilai parameter setiap percobaan

Skenario	n_estimator	max_depth	min_samples_split	RMSE
1	100	10	8	1.9273
2	200	10	8	1.9101
3	300	10	8	1.9086
4	400	10	8	1.9071
5	500	10	8	1.9098
6	600	10	8	1.9135
7	700	10	8	1.9133
8	800	10	8	1.9206
9	900	10	8	1.9196
10	1000	10	8	1.9163
11	100	5	8	1.9273
12	100	15	8	1.9273
13	100	20	8	1.9273
14	100	25	8	1.9273
15	100	30	8	1.9273
16	100	35	8	1.9273
17	100	40	8	1.9273
18	100	60	8	1.9273
19	100	80	8	1.9273
20	100	100	8	1.9273
21	100	10	4	1.9392
22	100	10	12	1.961
23	100	10	16	1.9606
24	100	10	20	2.0009
25	100	10	24	2.0055
26	100	10	32	2.0041

27	100	10	40	1.99
28	100	10	52	2.1471
29	100	10	68	2.1493
30	100	10	80	2.1493

Halaman ini sengaja dikosongkan