

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

PREDIKSI PENAMBAHAN JUMLAH TERKONFIRMASI POSITIF COVID-19 BERBASIS SOSIAL, EKONOMI DAN DEMOGRAFI MENGGUNAKAN *FBPROPHET* (STUDI KASUS DI JAWA TIMUR)

SITI NUR HASANAH



**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2021**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



PERNYATAAN MENGENAI SKRIPSI DAN SUMBER INFORMASI SERTA PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa skripsi dengan judul “Prediksi Penambahan Jumlah Terkonfirmasi Positif COVID-19 Berbasis Sosial, Ekonomi dan Demografi Menggunakan *FBProphet* (Studi Kasus di Jawa Timur)” adalah karya saya dengan arahan dari dosen pembimbing dan belum diajukan dalam bentuk apa pun kepada perguruan tinggi mana pun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka di bagian akhir skripsi ini.

Dengan ini saya melimpahkan hak cipta dari karya tulis saya kepada Institut Pertanian Bogor.

Bogor, Juli 2021

Siti Nur Hasanah
G64170036



ABSTRAK

SITI NUR HASANAH. Prediksi Penambahan Jumlah Terkonfirmasi Positif COVID-19 Berbasis Sosial, Ekonomi dan Demografi Menggunakan *FBProphet* (Studi Kasus di Jawa Timur). Dibimbing oleh YENI HERDIYENI dan MEDRIA KUSUMA DEWI HARDHIENATA.

Coronavirus Disease 2019 (COVID-19) merupakan penyakit menular yang ditetapkan sebagai pandemi di tahun 2020. Jawa Timur merupakan salah satu provinsi di Indonesia dengan jumlah terkonfirmasi positif COVID-19 paling tinggi. Jumlah terkonfirmasi positif COVID-19 yang berbeda di setiap daerah disebabkan oleh berbagai faktor, termasuk faktor sosial, ekonomi dan demografi daerah tersebut. Penelitian ini melakukan analisis *cluster* kabupaten/kota di Jawa Timur untuk mengetahui kondisi sosial, ekonomi dan demografi daerah serta analisis regresi untuk mengetahui kaitan faktor sosial, ekonomi dan demografi terhadap jumlah terkonfirmasi positif COVID-19. Selain itu pada penelitian ini juga dilakukan analisis prediktif penambahan jumlah terkonfirmasi positif COVID-19 menggunakan *FBProphet*. Berdasarkan hasil penelitian, kabupaten/kota di Jawa Timur terbagi menjadi lima *cluster* sesuai dengan karakteristik sosial, ekonomi dan demografinya. Jumlah kasus positif COVID-19 yang tinggi terdapat di *cluster* dengan nilai Produk Domestik Regional Bruto (PDRB) dan tingkat kepadatan penduduk yang tinggi. Hasil analisis regresi dengan nilai *adjusted r-squared* 0,86 menunjukkan bahwa atribut PDRB merupakan atribut yang berpengaruh signifikan terhadap jumlah terkonfirmasi positif COVID-19 di Jawa Timur. Penelitian ini memperkirakan jumlah terkonfirmasi positif COVID-19 masih terus naik hingga 30 Juli 2021. Model ini mampu memprediksi dengan nilai RMSLE 0,25.

Kata kunci: COVID-19, *FBProphet*, Sosial-ekonomi-demografi



ABSTRACT

SITI NUR HASANAH. *Prediction of Daily Positive Cases of COVID-19 Based on Socioeconomic and Demographics Using FBProphet (Case Study in East Java)*. Supervised by YENI HERDIYENI and MEDRIA KUSIMA DEWI HARDHIENATA.

Coronavirus Disease 2019 (COVID-19) is an infectious disease designated as a pandemic in 2020. East Java is one of the provinces in Indonesia with the highest number of positive cases of COVID-19. The number of positive COVID-19 case differences in each region are caused by various factors, including the area's social, economic, and demographic. This study conducted a cluster analysis of districts/cities in East Java to determine the social, economic, and demographic conditions of the area and regression analysis to determine the relationship of social, economic, and demographic factors to the number of positive COVID-19. In addition, this study also carried out a predictive analysis of increasing the number of confirmed positive COVID-19 using FBProphet. Based on the research results, districts/cities in East Java are divided into five clusters according to their social, economic, and demographic characteristics. Many positive cases of COVID-19 are found in clusters with a high value of Gross Regional Domestic Product (GDP regional) and population density. The results of the regression analysis with an adjusted r-squared value of 0.86 indicate that GDP attribute is an attribute that has a significant effect on the number of positive confirmed cases of COVID-19 in East Java. This study estimates that the addition of the positive number of COVID-19 will continue to rise until July 30, 2021. This model can predict with RMSLE value of 0.25.

Keywords: COVID-19, FBProphet, Socio-economic-demographic



**PREDIKSI PENAMBAHAN JUMLAH TERKONFIRMASI
POSITIF COVID-19 BERBASIS SOSIAL, EKONOMI DAN
DEMOGRAFI MENGGUNAKAN *FBPROPHET* (STUDI
KASUS DI JAWA TIMUR)**

SITI NUR HASANAH

Skripsi
sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer
pada
Departemen Ilmu Komputer

**DEPARTEMEN ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT PERTANIAN BOGOR
BOGOR
2021**



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Tim Penguji pada Ujian Skripsi:
Dr. Musthofa, S.Kom, Msc



@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Judul Skripsi : Prediksi Penambahan Jumlah Terkonfirmasi Positif COVID-19
Berbasis Sosial, Ekonomi dan Demografi Menggunakan
FBProphet (Studi Kasus di Jawa Timur)


Nama : Siti Nur Hasanah
NIM : G64170036

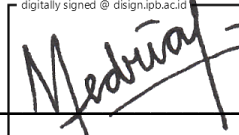
@Hak cipta milik IPB University

Disetujui oleh

Pembimbing 1:
Dr. Yeni Herdiyeni, S.Si, M.Kom

Pembimbing 2:
Dr. Medria Kusuma Dewi Hardhienata, S.Komp

digitally signed @ design.ipb.ac.id

0D625CA1-CE41-4DCD-BB8C-8D2EEB563A57

digitally signed @ design.ipb.ac.id

0D625CA1-CE41-4DCD-BB8C-8D2EEB563A57

Diketahui oleh

Ketua Departemen Ilmu Komputer:
Dr. Sony Hartono Wijaya, S.Kom, M.Kom
NIP. 19810809 200812 1 002

digitally signed

design.ipb.ac.id

Tanggal Ujian:
06 Juli 2021

Tanggal Lulus:

PRAKATA

Puji dan syukur penulis panjatkan kepada Allah Subhanahu wa Ta'ala atas segala karunia-Nya sehingga skripsi ini berhasil diselesaikan. Judul penelitian yang dilaksanakan sejak bulan Desember 2020 adalah Prediksi Penambahan Jumlah Terkonfirmasi Positif COVID-19 Berbasis Sosial, Ekonomi dan Demografi Menggunakan *FBProphet* (Studi Kasus di Jawa Timur). Penulis menyadari masih banyak kekurangan pada pelaksanaan penelitian dan penulisan skripsi ini. Pelaksanaan penelitian dan penulisan skripsi ini tidak dapat diselesaikan tanpa bantuan, bimbingan, serta dukungan dari berbagai pihak. Karena itu, penulis ingin mengucapkan terima kasih kepada:

1. Ibu, Ayah, serta seluruh keluarga penulis atas segala do'a dan kasih sayangnya
2. Ibu Dr. Yeni Herdiyeni, S.Si, M.Kom dan Ibu Dr. Medria K.D. Hardhienata, S.Komp selaku dosen pembimbing yang telah banyak memberi saran dan membimbing penulis sehingga dapat menyelesaikan penelitian dan penulisan skripsi ini.
3. Bapak Dr. Musthofa, S.Kom, Msc selaku dosen penguji penelitian ini yang juga banyak memberikan masukan selama proses sidang.
4. Teman-teman Ilkomerz 54 yang telah membantu proses perkuliahan di Ilmu Komputer.
5. Pihak-pihak lain yang selalu mendukung dan menemani penulis selama di IPB hingga memperoleh gelar sarjana.

Semoga skripsi ini bermanfaat dan menambah wawasan bagi pembaca.

Bogor, Juli 2021

Siti Nur Hasanah



DAFTAR ISI

DAFTAR TABEL	xi
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Manfaat	2
1.5 Ruang Lingkup	2
TINJAUAN PUSTAKA	3
2.1 <i>Coronavirus Disease 2019 (COVID-19)</i>	3
2.2 COVID-19 di Jawa Timur	3
2.3 Sosial, Ekonomi dan Demografi Provinsi Jawa Timur	3
2.4 <i>Facebook Prophet</i>	4
2.5 Validasi Model	6
2.6 <i>Agglomerative Hierarchical Clustering</i>	7
III METODE	8
3.1 Data Penelitian	8
3.2 Tahapan Penelitian	9
3.3 Lingkungan Pengembangan	15
IV HASIL DAN PEMBAHASAN	15
4.1 Eksplorasi dan Praproses Data	15
4.2 Analisis Kaitan Faktor Sosial, Ekonomi dan Demografi Terhadap Jumlah Kasus COVID-19	18
4.3 Prediksi Penambahan Jumlah terkonfirmasi positif COVID-19 di Jawa Timur	27
V SIMPULAN DAN SARAN	33
5.1 Simpulan	33
5.2 Saran	34
DAFTAR PUSTAKA	35
LAMPIRAN	37
RIWAYAT HIDUP	47

DAFTAR TABEL

1	Atribut sosial, ekonomi dan demografi pada penelitian	8
2	Statistik deskriptif 12 atribut sosial, ekonomi dan demografi	17
3	Contoh data sebelum dilakukan standarisasi	17
4	Data hasil standarisasi	18
5	Hasil <i>clustering</i> metode hirarki dengan pautan <i>Ward</i>	20
6	Rata-rata atribut setiap <i>cluster</i>	20
7	Kondisi COVID-19 di wilayah <i>cluster</i> nol	22
8	Kondisi COVID-19 di wilayah <i>cluster</i> satu	23
9	Kondisi COVID-19 di wilayah <i>cluster</i> dua	24
10	Kondisi COVID-19 di wilayah <i>cluster</i> tiga	25
11	Kondisi COVID-19 di wilayah <i>cluster</i> empat	26
12	Hasil analisis linear berganda	26
13	Koefisien determinasi	27
14	Daftar parameter yang dilakukan <i>tuning</i>	28
15	Perbandingan model prediksi	31
16	Hasil prediksi dan komponen <i>FBProphet</i>	31

DAFTAR GAMBAR

17	Ilustrasi proses <i>expanding window cross validation</i>	6
18	Tahapan penelitian	9
19	Tahapan analisis deskriptif kaitan faktor sosial, ekonomi dan demografi terhadap jumlah terkonfirmasi positif COVID-19	11
20	Tahapan analisis prediktif positif COVID-19 di Jawa Timur	14
21	Plot penambahan positif harian COVID-19 di Jawa Timur	15
22	Plot jumlah kumulatif positif COVID-19 per kabupaten/kota di Jawa Timur	16
23	Grafik kategori kasus COVID-19 di setiap kabupaten/kota	16
24	Koordinat paralel rata-rata atribut setiap <i>cluster</i>	21
25	Plot distribusi penambahan positif harian COVID-19 sebelum ditransformasi	28
26	Plot distribusi data setelah ditransformasi	28
27	Plot prediksi menggunakan model <i>default FBProphet</i>	29
28	Plot prediksi menggunakan model <i>FBProphet hyperparameter tuning</i>	30
29	Plot prediksi penambahan positif harian COVID-19 hingga 30 Juli 2021	32
30	Plot komponen <i>trend</i> pada <i>FBProphet</i>	32
31	Plot komponen <i>holiday</i> pada <i>FBProphet</i>	32
32	Plot komponen <i>seasonality</i> pada <i>FBProphet</i>	33

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



33	Tabel data sebelum standarisasi	37
34	<i>Heatmap</i> koefisien korelasi antar atribut	39
35	Nilai <i>p-value</i> korelasi antar atribut	39
36	Nilai <i>Variance Inflation Factor</i>	40
37	Jarak <i>Euclidean</i> antar kabupaten/kota di Jawa Timur berdasarkan data sosial, ekonomi dan demografi	41
38	Matriks pautan pada metode hirarki dengan pautan <i>Ward</i>	45
39	Dendogram hasil pengelompokan menggunakan metode <i>Agglomerative Hierarchical Clustering</i>	46

I PENDAHULUAN

1.1 Latar Belakang

Coronavirus Disease 2019 (COVID-19) merupakan suatu penyakit menular yang disebabkan oleh infeksi virus *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2) (Shereen *et al.* 2020). Berawal dari kasus pertama yang dilaporkan di Wuhan, China pada akhir tahun 2019 dalam beberapa hari penyakit ini menyebar dengan cepat ke berbagai kota lainnya di China. Sampai dengan saat ini telah menginfeksi hampir seluruh negara di dunia (Li *et al.* 2020; Shereen *et al.* 2020). Pada tanggal 11 Maret 2020, *World Health Organization* (WHO) menetapkan COVID-19 sebagai pandemi global. Di Indonesia, kasus COVID-19 pertama dilaporkan oleh Kementerian Kesehatan Republik Indonesia pada tanggal 2 Maret 2020 dan menyebar dengan cepat hampir ke seluruh daerah. Salah satu daerah dengan jumlah terkonfirmasi positif COVID-19 cukup tinggi di Indonesia adalah provinsi Jawa Timur. Menurut data Kementerian Kesehatan Republik Indonesia, sampai dengan 4 Mei 2021 jumlah terkonfirmasi positif COVID-19 di Jawa Timur sebanyak 148.688 atau 8,8% dari kasus nasional. Hal ini menunjukkan bahwa penyebaran COVID-19 di Provinsi Jawa Timur masih cukup tinggi.

Beberapa penelitian menyatakan bahwa faktor sosial, ekonomi dan demografi suatu daerah memiliki pengaruh dalam membentuk pola penyebaran COVID-19. Penelitian yang dilakukan oleh Sannigrahi *et al.* (2020) menyatakan bahwa struktur sosial seperti penghasilan masyarakat dan tingkat kemiskinan, serta struktur demografi, seperti total populasi suatu daerah berperan penting dalam menentukan terjadinya kasus COVID-19. Penelitian lainnya oleh Tosepu *et al.* (2020) menjelaskan beberapa faktor yang menyebabkan tingginya transmisi COVID-19 yaitu cuaca, tingginya mobilitas masyarakat, dan jumlah penduduk yang sangat padat. Di China Qiu *et al.* (2020) melakukan analisis mengenai aktivitas ekonomi masyarakat, hasilnya daerah dengan Produk Domestik Bruto (PDB) per kapita yang lebih tinggi memiliki tingkat penyebaran yang juga tinggi. Hal ini berkaitan dengan peningkatan interaksi sosial seiring dengan aktivitas perekonomian. Adanya informasi mengenai karakteristik sosial, ekonomi dan demografi suatu daerah dapat digunakan sebagai landasan untuk melakukan prediksi penyebaran COVID-19 di wilayah tersebut.

Analisis prediktif COVID-19 merupakan bagian penting dari penanganan penyakit. Hasil prediksi yang baik dapat digunakan dalam melakukan upaya pencegahan penyebaran penyakit yang semakin meluas. Prediksi COVID-19 dan penyakit menular lainnya telah dilakukan menggunakan berbagai pendekatan seperti pendekatan epidemiologi, statistika dan *Machine Learning* (Ardabili *et al.* 2020). Menurut Shinde *et al.* (2020) teknik *Machine Learning* merupakan pendekatan yang banyak digunakan karena kemampuannya dalam memprediksi pola penyakit dari data deret waktu. Jika diamati data jumlah terkonfirmasi positif COVID-19 di Jawa Timur memiliki *trend* naik, tidak stasioner dan dipengaruhi efek musiman. Salah satu teknik *Machine Learning* yang dapat digunakan untuk memprediksi data deret waktu dengan karakteristik tersebut adalah *Facebook Prophet* (FBProphet). FBProphet merupakan teknik yang digunakan untuk memprediksi data deret waktu berdasarkan model aditif. Teknik ini memodelkan

komponen-komponen dari data deret waktu seperti komponen *trend*, efek musiman dan liburan untuk memprediksi data. *FBProphet* mampu mengenali perubahan *trend* pada data dan bekerja dengan baik pada data yang memiliki efek musiman (Aditya Satrio *et al.* 2021). Oleh karena itu, teknik ini banyak digunakan dalam memprediksi data pada berbagai bidang seperti bisnis, ekonomi, kesehatan, lingkungan, bahkan juga epidemi seperti COVID-19.

Penelitian yang dilakukan oleh Gupta *et al.* (2021) membandingkan teknik *Support Vector Machine*, *Linear Regression* dan *FBProphet* dalam meramalkan kasus COVID-19 di India untuk jangka pendek. Hasilnya *FBProphet* memberikan hasil yang lebih baik dengan nilai *error* lebih kecil dibandingkan teknik lainnya. Namun penelitian tersebut hanya terbatas pada pembentukan model prediksi, belum terdapat analisis mengenai kaitan faktor sosial, ekonomi dan demografi terhadap COVID-19. Pada penelitian ini akan dilakukan analisis kaitan faktor sosial, ekonomi dan demografi terhadap COVID-19 yang kemudian akan digunakan sebagai landasan dalam melakukan analisis prediktif penambahan jumlah terkonfirmasi positif COVID-19 di Jawa Timur menggunakan *FBProphet*. Dengan melakukan dua jenis analisis ini, harapannya dapat memberikan informasi lebih baik mengenai kondisi COVID-19 di Provinsi Jawa Timur.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini adalah:

1. Bagaimana melakukan analisis deskriptif kaitan faktor sosial, ekonomi dan demografi daerah di Jawa Timur terhadap jumlah terkonfirmasi positif COVID-19.
2. Bagaimana menerapkan arsitektur *FBProphet* dalam melakukan analisis prediktif penambahan jumlah terkonfirmasi positif COVID-19 di Provinsi Jawa Timur.

1.3 Tujuan

Tujuan penelitian ini adalah:

1. Menganalisis kaitan faktor sosial, ekonomi dan demografi daerah Jawa Timur terhadap jumlah terkonfirmasi positif COVID-19.
2. Menerapkan arsitektur *FBProphet* untuk memprediksi penambahan jumlah kasus COVID-19 di Jawa Timur.

1.4 Manfaat

Penelitian ini menghasilkan analisis deskriptif kaitan faktor sosial, ekonomi dan demografi Provinsi Jawa Timur terhadap COVID-19 serta analisis prediktif penambahan jumlah terkonfirmasi positif COVID-19 untuk beberapa waktu kedepan. Harapannya hasil analisis deskriptif dan prediktif ini dapat menjadi sumber informasi dalam melakukan pencegahan penyebaran penyakit yang semakin meluas khususnya di Provinsi Jawa Timur.

1.5 Ruang Lingkup

Lingkup dari penelitian ini yaitu:

- 1 Data yang digunakan adalah data jumlah terkonfirmasi positif COVID-19 di Jawa Timur, serta data informasi sosial, ekonomi dan demografi Provinsi Jawa Timur.
- 2 Penelitian ini berfokus melakukan analisis kaitan faktor sosial, ekonomi dan demografi terhadap jumlah terkonfirmasi positif COVID-19.
- 3 Model prediksi penambahan jumlah terkonfirmasi positif COVID-19 yang dikembangkan pada penelitian ini khusus untuk Provinsi Jawa Timur.

II TINJAUAN PUSTAKA

2.1 *Coronavirus Disease 2019 (COVID-19)*

Penyakit virus corona (COVID-19) adalah suatu penyakit menular yang mempengaruhi sistem pernafasan manusia. Penyakit ini disebabkan oleh virus *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2) yang bersifat sangat mudah menular dan patogenik. COVID-19 ditandai dengan gejala demam, kelelahan, dan batuk kering, pada kondisi yang lebih parah COVID-19 dapat menyebabkan pneumonia, sindrom pernapasan akut, gagal ginjal dan bahkan kematian (Shinde *et al.* 2020; Tosepu *et al.* 2020). *Coronavirus* yang menyebabkan pandemi COVID-19 berasal dari kelompok yang sama dengan virus penyebab pandemi SARS 2002 dan MERS 2012, tetapi laju penularan COVID-19 paling tinggi dibandingkan dua pandemi sebelumnya (Shereen *et al.* 2020). Hal ini disebabkan karena sifat virusnya yang sangat menular dan penularan yang terjadi antar manusia. Penularan virus antar manusia dapat disebabkan karena terjadinya kontak dekat dengan seseorang yang terinfeksi dan penularan melalui droplet atau cairan yang dikeluarkan saat seseorang bersin, batuk dan berbicara (Shereen *et al.* 2020).

2.2 COVID-19 di Jawa Timur

Jawa Timur merupakan salah satu provinsi dengan jumlah COVID-19 cukup tinggi di Indonesia. Sampai dengan 4 Mei 2021 jumlah kasus terkonfirmasi COVID-19 di Jawa Timur sebanyak 148.688 atau 8,8% dari kasus nasional yang menempatkan Jawa Timur sebagai provinsi dengan jumlah kasus COVID-19 terbanyak keempat di Indonesia (Satgas Covid-19 2021). Kota Surabaya adalah daerah dengan jumlah terkonfirmasi positif terbanyak di Provinsi Jawa Timur dengan total positif 23.561 atau mencakup 15,85% keseluruhan kasus terkonfirmasi positif di Jawa Timur.

2.3 Sosial, Ekonomi dan Demografi Provinsi Jawa Timur

Secara ekonomi, Provinsi Jawa Timur merupakan penghubung antara kawasan Timur dan Barat Indonesia, khususnya sebagai pintu gerbang perdagangan antar pulau dan antar daerah. Pada aspek demografi Provinsi Jawa Timur memiliki tingkat kepadatan penduduk yang cukup tinggi. Berdasarkan data BPS Jawa Timur, jumlah penduduk di tahun 2020 yaitu sebanyak 40.665.696 jiwa dengan kepadatan penduduk per kilometer persegi sebesar 851 jiwa/km². Sebanyak 71,65% penduduk provinsi ini berada pada usia produktif (15 - 64 tahun). Provinsi ini memiliki nilai

rasio jenis kelamin 99.6 yang berarti bahwa terdapat 99 hingga 100 laki-laki untuk setiap 100 perempuan. Hal ini menunjukkan bahwa jumlah penduduk laki-laki lebih sedikit dibandingkan jumlah penduduk perempuan.

Pada sektor perekonomian, Provinsi Jawa Timur merupakan salah satu provinsi yang menyumbang angka besar pada nilai PDB Nasional. Pada tahun 2018 Provinsi Jawa Timur menyumbang 14,67% terhadap PDB Nasional. Pada tahun 2019, laju pertumbuhan ekonomi Provinsi Jawa Timur yaitu sebesar 5% yang melampaui pertumbuhan ekonomi Indonesia yang hanya sebesar 3,82%. Namun, pada tahun 2020 hampir seluruh provinsi di Indonesia mengalami penurunan termasuk Jawa Timur, provinsi ini mengalami penurunan sebesar 2,84% (BPS 2021).

2.4 Facebook Prophet

Facebook Prophet atau *FBProphet* adalah suatu teknik peramalan data yang dikembangkan oleh tim *Data Science Facebook* pada tahun 2017. Teknik ini menggunakan model deret waktu yang dapat diuraikan dengan tiga komponen utama yaitu: *trend*, *seasonality* (musim) dan *holiday* (hari libur). Model ini diformulasikan menggunakan persamaan berikut (Taylor dan Letham 2018):

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (1)$$

$g(t)$ adalah fungsi *trend* yang memodelkan perubahan non-periodik dalam nilai deret waktu, $s(t)$ mewakili perubahan periodik (misalnya harian, mingguan atau tahunan), $h(t)$ mewakili efek hari libur terhadap data, dan ϵ_t adalah perubahan lainnya yang tidak diakomodasi oleh model. Spesifikasi pada *FBProphet* ini mirip dengan model *Generalized Additive Model* (GAM), yaitu suatu model regresi yang digunakan untuk mengakomodasi pengaruh nonlinear dari peubah penjelas. *FBProphet* hanya menggunakan variabel waktu sebagai *regressor*, akan tetapi masih memungkinkan untuk menambah beberapa fungsi linear dan nonlinear dari waktu sebagai komponen (Taylor dan Letham 2018). Berikut adalah penjelasan untuk komponen utama dari *FBProphet*:

1. Fungsi *Trend*

Trend adalah komponen jangka panjang yang mendasari pertumbuhan atau penurunan pada data deret waktu. *Trend* menggambarkan pergerakan data sedikit demi sedikit menaik atau menurun. Ada dua model *trend* yang terdapat pada *FBProphet* yaitu *nonlinear saturating growth* dan *linear trend with changepoints*.

- Nonlinear, Saturating Growth*: model ini digunakan apabila terdapat asumsi bahwa pada saat meramalkan pertumbuhan data deret waktu, ada titik maksimum yang dapat dicapai, yang disebut dengan *capacity* dan peramalan harus mencapai titik ini. Persamaan dasar dari model logistik yaitu (Taylor dan Letham 2018):

$$g(t) = \frac{C}{1 + \exp(-k(t - m))} \quad (2)$$

Dengan C adalah *capacity*, k adalah laju pertumbuhan dan m adalah *offset* parameter.

Ketika melakukan peramalan pada data deret waktu, nilai *capacity* dan laju pertumbuhan tidak selalu konstan. Sehingga pada *FBProphet capacity* C didefinisikan sebagai $C(t)$ yaitu nilai *capacity* yang dapat berubah pada waktu t . Untuk melakukan penyesuaian pada nilai laju pertumbuhan, pada *FBProphet* didefinisikan parameter *changepoint* yang merepresentasikan titik perubahan di mana laju pertumbuhan pada data diizinkan untuk berubah. Sehingga persamaan *piecewise logistic growth* pada *FBProphet* adalah sebagai berikut (Taylor dan Letham 2018):

$$g(t) = \frac{C(t)}{1 + \exp(-(k + a(t)^T \delta)(t - (m + a(t)^T \gamma)))} \quad (3)$$

Dimana k adalah laju pertumbuhan yang disesuaikan dengan parameter $a(t)^T \delta$. Karena laju pertumbuhan k mengalami penyesuaian sehingga *offset* parameter m juga harus disesuaikan dengan vektor $a(t)^T \gamma$.

- b. *Linear Trend-With Changepoints*: model ini digunakan apabila tidak diketahui titik maksimum tertentu yang harus dicapai data, atau pada kondisi fungsi pertumbuhan linier atau konstan. Pada model ini data dibagi menjadi beberapa wilayah yang berbeda berdasarkan nilai *breakpoint* nya. Model ini menggunakan persamaan (Taylor dan Letham 2018):

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma) \quad (4)$$

Dimana k adalah laju pertumbuhan, δ adalah tingkat penyesuaian, m adalah *offset* parameter, γ adalah parameter penyesuaian yang membuat fungsi kontinu.

2. Seasonality

Seasonality yaitu pola data yang berulang pada kurun waktu tertentu. Pada *FBProphet* terdapat tiga jenis efek *seasonality* yaitu *daily*, *weekly* dan *yearly*. Namun, dapat juga ditambahkan efek *seasonality* lainnya sesuai dengan kebutuhan analisis. *Seasonality* pada *FBProphet* diperkirakan menggunakan persamaan Fourier berikut (Taylor dan Letham 2018):

$$s(t) = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi n t}{P}\right) + b_n \sin\left(\frac{2\pi n t}{P}\right) \right) \quad (5)$$

Dengan P adalah periode berulang pada data deret waktu (misalnya $P=365,25$ untuk pola tahunan atau $P=7$ untuk data mingguan). Dalam melakukan penyesuaian efek musiman pada data, maka perlu dilakukan estimasi terhadap $2N$ parameter $\beta = [a_1, b_1, \dots, a_N, b_N]^T$. Hal ini dilakukan dengan membangun matriks vektor musiman untuk setiap nilai yang sering ada pada data, misalnya dengan *seasonality yearly* dan $N=10$ maka:

$$X(t) = \left[\cos\left(\frac{2\pi(1)t}{365,25}\right), \dots, \sin\left(\frac{2\pi(10)t}{365,25}\right) \right]$$

Sehingga persamaan akhir pada komponen seasonality adalah:

$$s(t) = X(t)\beta \quad (6)$$

3. *Holidays and Events*

Liburan dan peristiwa tertentu dapat menyebabkan terjadinya perubahan yang dapat diprediksi pada data. Misalnya, pada Hari Raya maka sebagian besar masyarakat akan melakukan perjalanan untuk mengunjungi keluarga. Selama periode ini, maka kemungkinan interaksi secara langsung lebih banyak terjadi. Adanya hari libur ini dapat mempengaruhi jumlah kasus COVID-19 di suatu daerah. Pada *FBProphet* efek liburan dan peristiwa tertentu dapat ditambahkan oleh peneliti. Untuk setiap *holiday i*, misalkan D_i adalah himpunan tanggal masa lalu dan masa depan untuk *holiday i*. Fungsi indikator ditambahkan yang mewakili waktu t selama *holiday i*, selain itu ditambahkan juga parameter k pada setiap hari libur yang merupakan nilai perubahan yang sesuai pada data. Hal ini dilakukan mirip dengan *seasonality* dengan membuat sebuah matriks regresi berikut (Taylor dan Letham 2018):

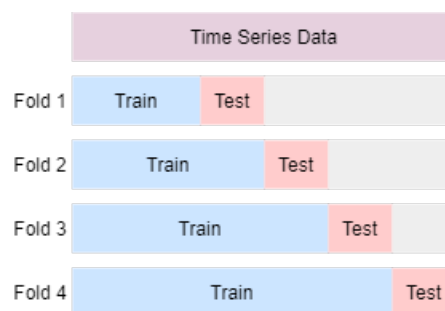
$$Z(t) = [1(t \in D_1), \dots, 1(t \in D_L)] \quad (7)$$

persamaan akhir komponen *holiday* adalah:

$$h(t) = Z(t)k \quad (8)$$

2.5 Validasi Model

Pada penelitian ini untuk melakukan validasi dan evaluasi model, maka digunakan teknik *expanding window cross validation*. Pada *FBProphet* hal ini dilakukan dengan memilih titik potong pada data, kemudian untuk setiap titik tersebut akan dilakukan penyesuaian sehingga data yang digunakan untuk pelatihan hanya sampai dengan titik potong tersebut. Teknik *cross validation* diterapkan untuk menghindari terjadinya *overfitting*. Teknik ini bekerja mirip dengan teknik *k-fold cross validation* biasanya, data dibagi menjadi beberapa bagian (*fold*) pada periode waktu yang berbeda, ukuran data *train* akan bertambah pada setiap *fold* (Germaine 2019). Proses ini diilustrasikan pada Gambar 1.



Gambar 1 Ilustrasi proses *expanding window cross validation*

Performa model pada *cross validation* diukur dengan menghitung nilai *Root Mean Squared Log Error* (RMSLE). RMSLE adalah salah satu metrik evaluasi yang digunakan untuk menilai kebaikan suatu model. RMSLE mempertimbangkan

kesalahan relatif antara data yang diprediksi dan nilai aktual. RMSLE akan memberikan nilai penalti yang lebih besar untuk prediksi yang terlalu rendah dari variabel aktual (Saxena 2019). Persamaan untuk menghitung nilai RMSLE adalah:

$$RMSLE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(p_i + 1) - \log(a_i + 1))^2} \quad (9)$$

dimana:

- n = panjang dari data deret waktu
- p_i = nilai prediksi saat ke- i
- a_i = nilai aktual saat ke- i

2.6 Agglomerative Hierarchical Clustering

Agglomerative Hierarchical Clustering adalah salah satu teknik pengelompokan data yang membentuk tingkatan atau hirarki dari data. Teknik ini dimulai dengan menganggap bahwa setiap objek adalah satu *cluster*, kemudian dua objek dengan jarak terdekat akan bergabung menjadi satu *cluster*. Proses ini akan berlanjut hingga terbentuk satu *cluster* yang terdiri dari keseluruhan objek. Untuk mengukur kedekatan antar objek digunakan metrik *Euclidean Distance* dengan persamaan:

$$d_{ij} = \sqrt{\sum_{k=1}^N (X_{ik} - X_{jk})^2} \quad (10)$$

Dimana:

- d_{ij} = jarak antara objek ke- i dengan objek ke- j
- N = jumlah variabel *cluster*
- X_{ik} = data ke- i pada variabel ke- k
- X_{jk} = data ke- j pada variabel ke- k

Ada berbagai teknik pautan yang dapat digunakan dalam *hierarchical clustering*, salah satunya adalah *Ward's Linkage*. Pada *Ward's Linkage* nilai *error sum of squares* (SSE) dihitung untuk menentukan objek yang mana yang akan digabungkan pada langkah berikutnya. SSE didefinisikan dengan persamaan:

$$SSE = \sum_{j=1}^p \left(\sum_{i=1}^N (x_{ij} - \bar{x}_j)^2 \right) \quad (11)$$

Dimana:

- x_{ij} = Nilai variabel ke- j dari objek ke- i pada suatu *cluster*
- \bar{x}_j = rata-rata nilai variabel ke- j pada suatu *cluster*
- p = banyaknya variabel yang diukur
- N = jumlah objek pada *cluster* yang terbentuk

Untuk melakukan penggabungan dua *cluster*, misal *cluster* P dan Q dilakukan dengan meminimalkan peningkatan SSE yang didefinisikan sebagai jarak antara *cluster* P dan *cluster* Q seperti berikut:

$$I_{(PQ)} = SSE_{PQ} - (SSE_P + SSE_Q) \quad (12)$$

Dengan kata lain, metode *Ward* menghitung jarak antara anggota *cluster* dan pusat *cluster*. Pusat dari sebuah *cluster* didefinisikan sebagai titik di mana jumlah kuadrat jarak *Euclidean* antara titik itu sendiri dan yang lainnya di dalam *cluster* minimum.

III METODE

3.1 Data Penelitian

Penelitian ini menggunakan data kasus terkonfirmasi positif COVID-19 di Jawa Timur dari tanggal 21 April 2020 sampai dengan 30 Juni 2021. Selain itu digunakan juga data sosial, ekonomi dan demografi Provinsi Jawa Timur tahun 2020. Tabel 1 berisi informasi mengenai dua belas atribut sosial, ekonomi dan demografi yang digunakan pada penelitian ini.

Tabel 1 Atribut sosial, ekonomi dan demografi pada penelitian

Nama Atribut	Tipe Data	Deskripsi
Indeks Pembangunan Manusia (IPM)	Numerik	IPM merupakan indikator yang digunakan untuk mengukur keberhasilan dalam upaya membangun kualitas hidup manusia (masyarakat/penduduk). IPM dibentuk oleh 3 dimensi dasar, yaitu umur panjang dan hidup sehat, pengetahuan dan standar hidup layak. Penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran per kapita per bulan yang berada di bawah garis kemiskinan
Persentase Penduduk Miskin	Numerik	Penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran per kapita per bulan yang berada di bawah garis kemiskinan
Jumlah Rumah Sakit Umum	Numerik	Jumlah rumah sakit yang memberikan pelayanan kesehatan pada semua bidang dan jenis penyakit.
Jumlah Dokter	Numerik	Jumlah dokter per kabupaten/kota di Jawa Timur.
Produk Domestik Regional Bruto-Atas Dasar Harga Konstan (PDRB-ADHK)	Numerik	PDRB-ADHK menunjukkan nilai tambah barang dan jasa yang dihitung menggunakan harga yang berlaku pada satu tahun tertentu sebagai dasar. Biasanya digunakan untuk mengetahui pertumbuhan ekonomi dari tahun ke tahun
Pengeluaran per Kapita	Numerik	Biaya yang dikeluarkan untuk konsumsi semua anggota rumah tangga selama sebulan dibagi dengan banyaknya anggota rumah tangga

Tingkat Pengangguran Terbuka (TPT)

Numerik

yang telah disesuaikan dengan paritas daya beli

Persentase jumlah pengangguran terhadap jumlah angkatan kerja. Mengindikasikan besarnya persentase angkatan kerja yang termasuk dalam pengangguran. TPT yang tinggi menunjukkan bahwa terdapat banyak angkatan kerja yang tidak terserap pada pasar kerja.

Rasio Kelamin

Jenis

Numerik

Perbandingan antara jumlah penduduk pria dan jumlah penduduk wanita pada suatu daerah dan pada waktu tertentu, yang biasanya dinyatakan dalam banyaknya penduduk pria per 100 wanita.

Jumlah Populasi

Numerik

Total orang yang sudah menetap di suatu wilayah paling sedikit enam bulan atau kurang dari enam bulan tetapi bermaksud untuk menetap.

Kepadatan Penduduk

Numerik

Rasio banyaknya penduduk per kilometer persegi.

Persentase Penduduk

Numerik

Perbandingan jumlah penduduk perkotaan (*urban population*) terhadap total seluruh penduduk yang ada pada daerah tersebut.

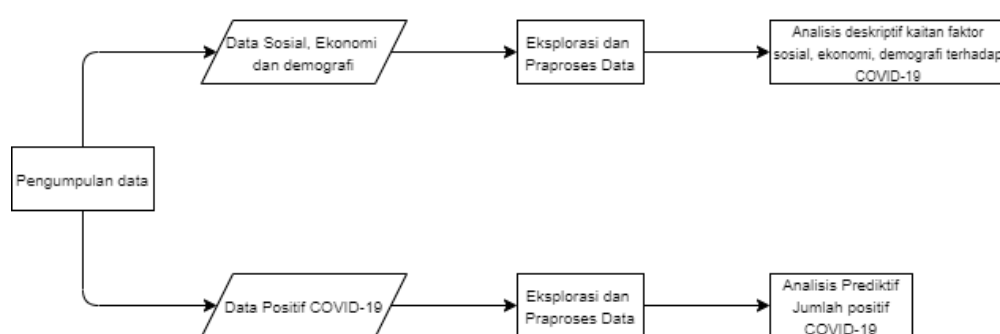
Persentase Penduduk Lansia

Numerik

Perbandingan jumlah penduduk usia 65 tahun keatas terhadap total keseluruhan penduduk pada daerah tersebut.

3.2 Tahapan Penelitian

Penelitian ini terdiri atas dari empat tahapan utama yaitu, pengumpulan dan pembersihan data, eksplorasi dan praproses data, analisis kaitan faktor sosial, ekonomi dan demografi terhadap jumlah kasus COVID-19, dan prediksi jumlah kasus positif COVID-19 di Jawa Timur. Tahapan penelitian dapat dilihat pada Gambar 2.



Gambar 2 Tahapan penelitian

3.2.1 Pengumpulan dan Pembersihan Data

Pada tahapan ini data terkonfirmasi positif COVID-19 diunduh dari situs Jatim Tanggap COVID (<http://infocovid19.jatimprov.go.id/>). Data yang diunduh adalah data dari tanggal 21 April 2020 sampai dengan tanggal 30 Juni 2021. Data yang bukan nama kabupaten/kota di Jawa Timur dihilangkan. Atribut yang disimpan hanya atribut terakhir update, kabupaten/kota, konfirmasi (hari ini), konfirmasi (penambahan) dan tingkat kematian. Sebanyak 465 fail yang sudah dibersihkan, kemudian digabungkan menjadi satu fail excel. Data sosial, ekonomi dan demografi diambil dari buku Provinsi Jawa Timur Dalam Angka 2021. Data yang dipilih hanya dua belas atribut yang sudah disebutkan pada subbab sebelumnya, kemudian data ini disalin ke dalam satu fail excel.

3.2.2 Eksplorasi dan Praproses Data

Tahapan eksplorasi data dilakukan untuk mendapatkan gambaran umum mengenai data yang akan digunakan. Pada tahapan ini dilihat bagaimana kondisi terkini dari kasus COVID-19 Jawa Timur. Tingkat keparahan kasus COVID-19 di setiap kabupaten/kota dibagi menjadi lima kategori yaitu sangat rendah, rendah, sedang, tinggi dan sangat tinggi berdasarkan kasus kumulatif positif COVID-19 pada tanggal 30 Juni 2021. Yang menjadi nilai batas pada kategorisasi adalah nilai rata-rata (\bar{x}) dan simpangan baku (S) kasus positif COVID-19. Proses kategorisasi dilakukan sebagai berikut:

- Sangat tinggi (ST). Kasus COVID-19 dikategorikan sangat tinggi jika:

$$x \geq \bar{x} + (1,5 \times S)$$

- Tinggi (T). Kasus COVID-19 dikategorikan tinggi jika:

$$\bar{x} + (0,5 \times S) \leq x < \bar{x} + (1,5 \times S)$$

- Sedang (S). Kasus COVID-19 dikategorikan sedang jika:

$$\bar{x} - (0,5 \times S) \leq x < \bar{x} + (0,5 \times S)$$

- Rendah (R). Kasus COVID-19 dikategorikan rendah jika:

$$\bar{x} - (1,5 \times S) \leq x < \bar{x} - (0,5 \times S)$$

- Sangat rendah (SR). Kasus COVID-19 dikategorikan sangat rendah jika:

$$x < \bar{x} - (1,5 \times S)$$

Pada data sosial, ekonomi dan demografi tahapan eksplorasi data dilakukan dengan melihat deskripsi data melalui statistika deskriptif setiap atribut. Setelah dilakukan eksplorasi data, berikutnya dilakukan tahapan praproses. Tujuan dari praproses data adalah agar data siap diolah dan dapat digunakan untuk proses berikutnya. Data sosial, ekonomi dan demografi memiliki satuan yang bervariasi sehingga data ditransformasi ke dalam bentuk normal baku $Z \sim N(0, 1)$. Persamaan untuk transformasi normal baku adalah:

$$Z_{ij} = \frac{x_{ij} - \bar{x}_i}{S_{xi}} \quad (13)$$

dengan:

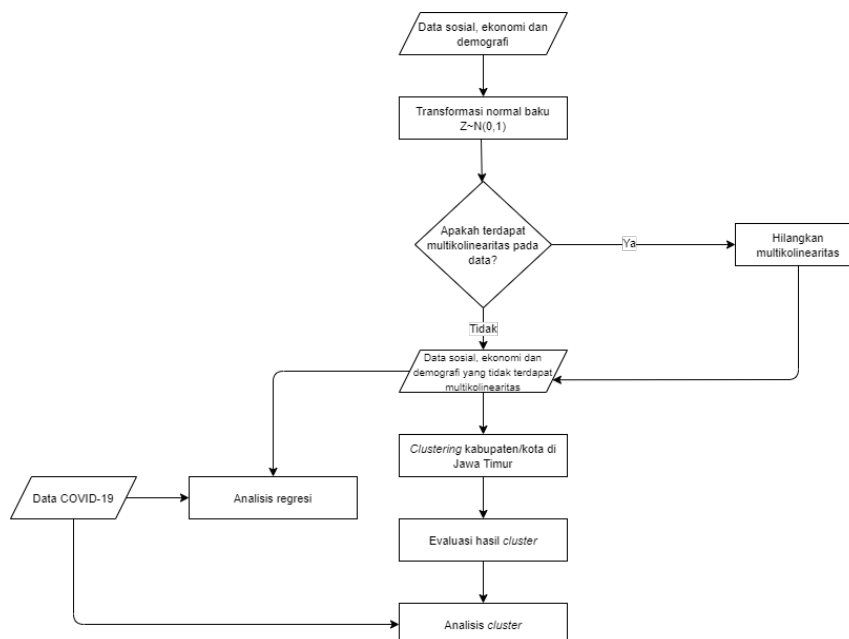
x_{ij} = data ke- j pada variabel ke- i

\bar{x}_i = rata-rata variabel ke- i

S_{xi} = simpangan baku variabel ke- i

Proses transformasi normal baku hanya dilakukan pada data sosial, ekonomi dan demografi.

3.2.3 Analisis Kaitan Faktor Sosial, Ekonomi dan Demografi Terhadap Jumlah Terkonfirmasi Positif COVID-19



Gambar 3 Tahapan analisis deskriptif kaitan faktor sosial, ekonomi dan demografi terhadap jumlah terkonfirmasi positif COVID-19

Gambar 3 menunjukkan tahapan analisis kaitan faktor sosial, ekonomi dan demografi terhadap jumlah kasus COVID-19 yang dilakukan pada penelitian ini yaitu:

1. Memeriksa multikolinearitas antar atribut pada data sosial, ekonomi dan demografi. Multikolinearitas adalah suatu kondisi dimana terdapat hubungan yang kuat antara dua variabel atau lebih. Pada analisis multivariat adanya multikolinearitas dapat memberikan dampak yang tidak baik (Ulinnuha dan Veriani 2020). Ada beberapa cara untuk memeriksa multikolinearitas di antaranya adalah dengan menghitung nilai korelasi *Pearson* antar variabel atau menghitung nilai *Variance Inflation Factor* (VIF). Pada penelitian ini untuk memeriksa adanya multikolinearitas, pertama dihitung nilai koefisien korelasi *Pearson* antar atribut. Koefisien korelasi *Pearson* mengukur kekuatan hubungan linear antara dua variabel. Nilai yang mendekati 1 atau -1 menunjukkan bahwa terdapat hubungan yang kuat antara kedua variabel tersebut. Persamaan untuk menghitung nilai korelasi *Pearson* yaitu:

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}} \quad (14)$$

dengan:

r = koefisien korelasi
 n = jumlah observasi
 X = atribut pertama
 Y = atribut kedua

Setelah menghitung nilai koefisien korelasi *Pearson*, berikutnya dilakukan perhitungan nilai *Variance Inflation Factor* (VIF) pada setiap atribut. VIF digunakan untuk mengukur seberapa besar kenaikan ragam dari dugaan antar variabel bebas (Shrestha 2020). VIF memberikan informasi mengenai seberapa baik variabel bebas dapat diprediksi menggunakan variabel bebas lainnya. Nilai VIF suatu variabel diperoleh dengan melakukan regresi antar variabel bebas. Misalkan terdapat 12 variabel bebas $X_1, X_2, X_3, \dots, X_{11}, X_{12}$. Maka untuk menghitung nilai VIF dari variabel X_1 , dilakukan *multiple linear regression* dengan variabel X_1 sebagai variabel terikat dan variabel lainnya sebagai variabel bebas. Dari proses regresi ini diperoleh nilai *r-squared*, yang kemudian digunakan untuk menghitung nilai VIF sesuai dengan formula berikut:

$$VIF = \frac{1}{1 - R^2} \quad (15)$$

dengan:

R^2 = nilai *R-squared* dari regresi atribut X terhadap atribut lainnya

Semakin besar nilai *r-squared* maka nilai VIF juga semakin meningkat. Nilai *r-squared* yang tinggi menunjukkan bahwa suatu variabel bebas dijelaskan dengan sangat baik oleh variabel bebas lainnya. Untuk menghitung nilai VIF dari variabel bebas lainnya dilakukan hal yang sama dengan perhitungan nilai VIF variabel X_1 . Nilai $VIF = 1$ menunjukkan bahwa variabel bebas tidak saling berkorelasi, jika $1 < VIF < 5$ berarti variabel berkorelasi sedang, dan jika nilai $5 \leq VIF < 10$ berarti terdapat korelasi yang kuat antar variabel bebas (Shrestha 2020). Salah satu cara yang dapat digunakan untuk menghilangkan multikolinearitas yaitu dengan tidak mengikutsertakan atribut yang saling berpengaruh tersebut pada proses analisis berikutnya (Ulinnuha dan Veriani 2020). Pada penelitian ini, dilakukan beberapa iterasi perhitungan VIF. Misalkan pada iterasi pertama, dari 12 variabel bebas, X_3 adalah salah satu variabel yang memiliki nilai $VIF > 5$ dan variabel ini memiliki korelasi yang tinggi dengan variabel lainnya. Maka X_3 akan *didrop*, kemudian pada iterasi kedua dilakukan perhitungan VIF, namun tidak mengikutsertakan variabel X_3 . Proses ini akan dilakukan hingga tidak ada lagi variabel yang memiliki nilai $VIF > 5$.

2. Melakukan *clustering* kabupaten/kota di Jawa Timur menjadi lima *cluster* dengan teknik *Agglomerative Hierarchical Clustering* berdasarkan data sosial, ekonomi dan demografi.
3. Melakukan evaluasi dari *cluster* yang terbentuk. *Cluster* yang terbentuk dievaluasi dengan melihat nilai koefisien *cophenetic cluster*. Koefisien *cophenetic* dihitung dengan persamaan berikut (Saraçlı et al. 2013):

$$c = \frac{\sum_{i < j} (x(i, j) - x)(t(i, j) - t)}{\sqrt{[\sum_{i < j} (x(i, j) - x)^2] [\sum_{i < j} (t(i, j) - t)^2]}} \quad (16)$$

dengan:

$x(i, j)$ = jarak *Euclidean* antara objek i dan j .
 $t(i, j)$ = jarak objek i dan j pada dendrogram
 x = rata-rata $x(i, j)$
 t = rata-rata $t(i, j)$

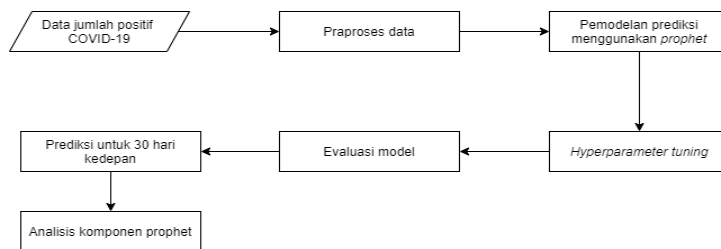
Nilai koefisien *cophenetic* yang baik adalah nilai yang mendekati 1.

4. Menganalisis karakteristik sosial, ekonomi dan demografi serta kondisi COVID-19 pada setiap *cluster*. Karakteristik sosial, ekonomi dan demografi dideskripsikan dengan melihat nilai rata-rata setiap atribut pada masing-masing *cluster*. Kemudian, nilai rata-rata ini juga dikategorikan menjadi sangat tinggi, tinggi, sedang, rendah, dan sangat rendah. Proses kategorisasi yang dilakukan sama dengan proses kategorisasi kasus COVID-19. Berikutnya, kondisi COVID-19 di setiap *cluster* diamati berdasarkan kategori jumlah terkonfirmasi positif dan tingkat kematian karena COVID-19 di masing-masing daerah anggota *cluster*.
5. Melakukan *multiple linear regression* untuk mengetahui atribut yang paling berpengaruh terhadap jumlah positif COVID-19. Pada analisis regresi ini, yang menjadi variabel terikat adalah jumlah kasus terkonfirmasi COVID-19 per kabupaten/kota pada tanggal 30 Juni 2021. Sedangkan data sosial, ekonomi dan demografi digunakan sebagai variabel bebas pada model regresi. Persamaan *multiple linear regression* yaitu (Basumatary dan Jain 2020):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{12} X_{12} + \epsilon \quad (17)$$

Dimana Y adalah variabel terikat yaitu jumlah terkonfirmasi positif COVID-19. X_1, X_2, \dots, X_{12} adalah variabel bebas yaitu atribut sosial, ekonomi dan demografi. $\beta_0, \beta_1, \dots, \beta_{12}$ adalah titik potong dan koefisien regresi dari setiap variabel, ϵ adalah nilai *error* pada model.

3.2.4 Peramalan Jumlah Kasus Positif COVID-19 di Jawa Timur Menggunakan *FBProphet*



Gambar 4 Tahapan analisis prediktif positif COVID-19 di Jawa Timur

Gambar 4 menunjukkan tahapan yang dilakukan untuk melakukan prediksi jumlah kasus positif COVID-19 di Provinsi Jawa Timur pada penelitian ini yaitu:

1. Memilih data yang akan digunakan sebagai input model prediksi. Karena yang ingin diprediksi adalah penambahan jumlah terkonfirmasi positif COVID-19, maka data yang digunakan adalah data konfirmasi (penambahan). Data ini masih terpisah per kabupaten/kota, sehingga harus dilakukan proses agregasi data.
2. Melakukan tahapan praproses data, yaitu penyesuaian pada data agar dapat digunakan sebagai input pada *FBprophet*. *FBProphet* membutuhkan sebuah *dataframe* yang terdiri dari dua kolom yaitu kolom ds dan kolom y. ds adalah kolom tanggal yang mengikuti format pandas *datetime* (tahun-bulan-tanggal), y adalah kolom yang berisi data yang ingin dilakukan prediksi. Dalam penelitian ini y merupakan data penambahan jumlah terkonfirmasi positif COVID-19 di Jawa Timur periode 21 April 2020 sampai dengan 30 Juni 2021. Kemudian pada tahapan praproses juga dilakukan transformasi data. Transformasi data dibutuhkan untuk membuat distribusi pada data mendekati distribusi normal. Persamaan yang digunakan pada transformasi Box-Cox pada data deret waktu y_t , dengan $t = 1, \dots, n$ adalah (Proietti dan Lutkepohl 2011):

$$y_t(\lambda) = \begin{cases} \frac{y_t^\lambda}{\lambda}, & \lambda \neq 0 \\ \ln y_t, & \lambda = 0 \end{cases} \quad (18)$$

3. Membuat model prediksi menggunakan *FBprophet*, dengan melakukan penyesuaian parameter *trend*, *seasonality* dan *holiday*.
4. Melakukan *hyperparameter tuning* dengan bantuan *grid search* untuk menemukan parameter terbaik.
5. Mengevaluasi model yang telah dibuat dengan menerapkan *expanding window cross validation*.
6. Melakukan prediksi jumlah terkonfirmasi positif dan penambahan positif COVID-19 di Jawa Timur untuk 30 hari berikutnya menggunakan model terbaik. Untuk memperoleh hasil aktual dari

prediksi, maka data yang sebelumnya ditransformasi harus di invers kembali.

7. Melakukan analisis komponen *trend*, *seasonality* dan *holiday* hasil pemodelan dengan *FBProphet*.

3.3 Lingkungan Pengembangan

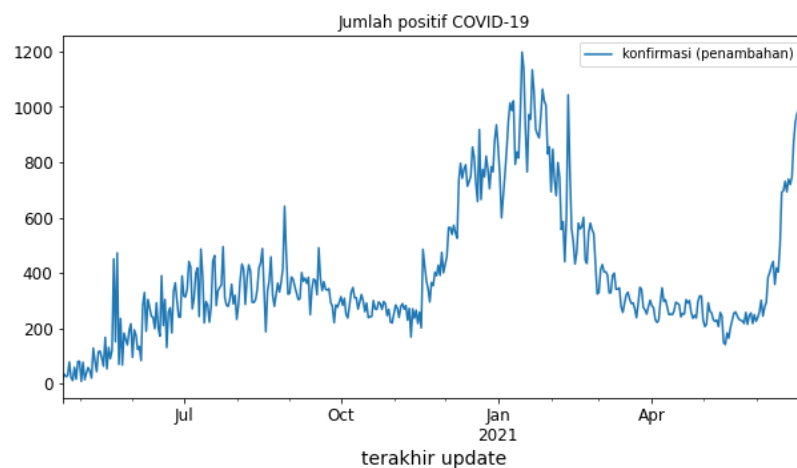
Spesifikasi perangkat keras dan perangkat lunak yang digunakan untuk penelitian ini adalah sebagai berikut:

1. Perangkat keras berupa laptop dengan spesifikasi:
 - Prosesor Intel Core i5-8265U CPU 1,6 GHz
 - RAM 8GB
 - NVIDIA GeForce MX250
2. Perangkat lunak yang digunakan, yaitu:
 - Sistem Operasi Windows 10 64-bit
 - Microsoft Excel 2016 untuk pengolahan data sederhana
 - Bahasa Pemrograman Python 3.8.5
 - Jupyter Notebook

IV HASIL DAN PEMBAHASAN

4.1 Eksplorasi dan Praproses Data

4.1.1 Deskripsi Data

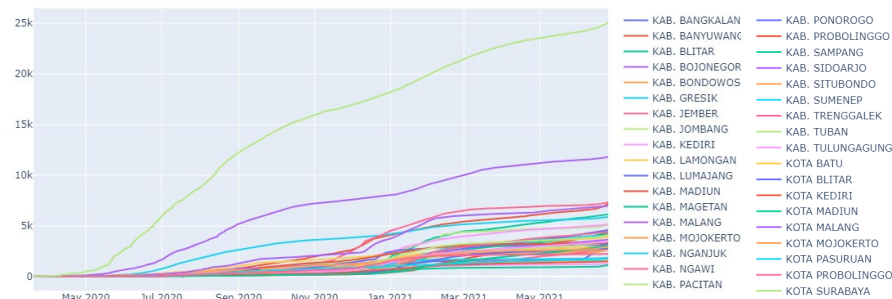


Gambar 5 Plot penambahan positif harian COVID-19 di Jawa Timur

Eksplorasi data dilakukan untuk memperoleh gambaran awal mengenai karakteristik data. Gambar 5 menunjukkan plot data penambahan positif COVID-19 yang berfluktuasi. Pada April 2020 sampai Februari 2021 data memiliki *trend* naik, kemudian pada Maret sampai Mei 2021 *trend* data mulai turun, namun sejak Juni 2021 *trend* data kembali naik. Simpangan baku penambahan positif COVID-19 di Jawa Timur yaitu 244,86. Hal tersebut menunjukkan bahwa data penambahan positif COVID-19 di Jawa Timur cukup

beragam dari waktu ke waktu. Penambahan positif COVID-19 tertinggi terjadi pada tanggal 15 Januari 2021 yaitu sebanyak 1198. Sedangkan penambahan kasus terendah terjadi pada 2 Mei 2020 sebanyak 6 kasus.

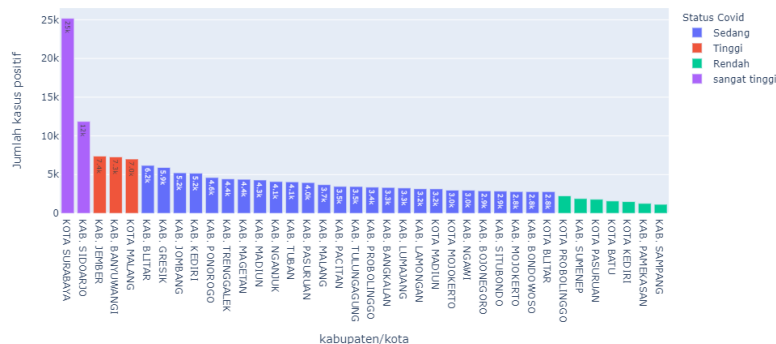
Jumlah Kasus Positif COVID-19



Gambar 6 Plot jumlah kumulatif positif COVID-19 per kabupaten/kota di Jawa Timur

Gambar 6 menunjukkan plot jumlah kumulatif positif COVID-19 per kabupaten/kota di Provinsi Jawa Timur. Di sebagian besar kabupaten/kota *trend* jumlah terkonfirmasi positif COVID-19 masih terus meningkat dari April 2020 sampai dengan Juni 2021. Kota Surabaya adalah daerah dengan jumlah serta pertambahan kasus paling tinggi di Provinsi Jawa Timur, diikuti Kabupaten Sidoarjo pada posisi kedua. Sedangkan Kabupaten Sampang adalah daerah dengan *trend* dan jumlah kasus positif paling rendah.

Jumlah kasus positif Covid-19 per kabupaten/kota



Gambar 7 Grafik kategori kasus COVID-19 di setiap kabupaten/kota

Grafik pada Gambar 7 menunjukkan kategori kondisi COVID-19 di setiap kabupaten/kota di Jawa Timur berdasarkan jumlah terkonfirmasi positif COVID-19 pada tanggal 30 Juni 2021. Kota Surabaya merupakan wilayah dengan jumlah terkonfirmasi positif COVID-19 paling tinggi dengan jumlah kasus sebanyak 25.179. Sedangkan daerah dengan jumlah terkonfirmasi positif COVID-19 paling rendah adalah Kabupaten Sampang yaitu sebanyak 1156 kasus. Terdapat perbedaan yang sangat besar antara daerah dengan jumlah kasus tertinggi dengan daerah lainnya di Jawa Timur. Hanya ada dua daerah yang termasuk kategori kasus positif COVID-19 sangat tinggi yaitu, Kota Surabaya

dan Kabupaten Sidoarjo. Sebagian besar daerah ada pada kategori sedang, yaitu sebanyak 26 kabupaten/kota. Tiga daerah ada pada kategori tinggi dan tujuh daerah lainnya termasuk pada kategori rendah. Tidak ada satupun daerah di Jawa Timur yang termasuk kategori sangat rendah.

Pada data sosial, ekonomi dan demografi, gambaran karakteristik data dilihat melalui statistik deskriptif dari setiap atribut sosial, ekonomi dan demografi kabupaten/kota.

Tabel 2 Statistik deskriptif 12 atribut sosial, ekonomi dan demografi

Atribut	Nilai minimum	Nilai maksimum	Rata-rata	Simpangan baku
Jumlah rumah sakit umum	2	37	7,974	6,828
Jumlah dokter	79	1.084	208,079	180,826
Indeks pembangunan manusia	62,7	82,23	71,871	5,046
Persentase penduduk miskin	3,89	22,78	11,021	4,568
PDRB	4.722,55	390.936,4	42.550,26	65.792,16
Total pengeluaran perkapita	667.972	1.879.145	1.054.290	294.458,8
Tingkat pengangguran terbuka	2,28	10,97	5,624	2,003
Kepadatan penduduk	295	8.200	1.922,711	2.107,439
Jumlah populasi	132.434	2.874.314	1.070.150	668.739,8
Persentase penduduk perkotaan	20,06	100	56,624	28,394
Rasio jenis kelamin	93,3	102,03	99,289	2,005
Persentase penduduk lansia	8,19	20,31	13,848	3,077

Tabel 2 menunjukkan beberapa nilai statistik deskriptif dari 12 atribut sosial, ekonomi dan demografi kabupaten/kota di Jawa Timur. Adanya perbedaan satuan antar atribut menyebabkan nilai rata-rata dan simpangan baku sangat beragam. Perbedaan nilai minimum dan maksimum yang cukup besar pada beberapa atribut menandakan bahwa belum merata dan seimbang kondisi sosial, ekonomi dan demografi di Jawa Timur.

4.1.2 Standarisasi Data

Data sosial, ekonomi dan demografi yang digunakan pada penelitian ini memiliki satuan yang berbeda, sehingga nilai antar atribut terpaut cukup jauh. Oleh karena itu, harus dilakukan proses standarisasi. Pada Tabel 3 ditampilkan contoh data sebelum dilakukan proses standarisasi dan pada Tabel 4 ditampilkan contoh data sesudah standarisasi. Data sosial, ekonomi dan demografi lengkap sebelum standarisasi terdapat pada Lampiran 1.

Tabel 3 Contoh data sebelum dilakukan standarisasi

Kabupaten/Kota	Jumlah rumah sakit umum	Jumlah dokter	...	Rasio jenis kelamin	Persentase penduduk lansia
Kab. Bangkalan	3	110	...	97,24	12,83
Kab. Banyuwangi	12	316	...	100,27	15,84
Kab. Blitar	8	184	...	101,53	17,11
⋮	⋮	⋮	⋮	⋮	⋮
Kota Pasuruan	2	96	...	100,08	9,46
Kota Probolinggo	2	131	...	98,35	10,66
Kota Surabaya	37	1084	...	98,35	9,16

Tabel 4 Data hasil standarisasi

Kabupaten/Kota	Jumlah rumah sakit umum	Jumlah dokter	...	Rasio jenis kelamin	Persentase penduduk lansia
Kab. Bangkalan	-0,738	-0,549	...	-1,036	-0,335
Kab. Banyuwangi	0,597	0,604	...	0,495	0,655
Kab. Blitar	0,004	-0,134	...	1,132	1,074
⋮	⋮	⋮	⋮	⋮	⋮
Kota Pasuruan	-0,886	-0,628	...	0,399	-1,445
Kota Probolinggo	-0,886	-0,431	...	-0,474	-1,050
Kota Surabaya	4,308	4,909	...	-0,474	-1,544

4.2 Analisis Kaitan Faktor Sosial, Ekonomi dan Demografi Terhadap Jumlah Kasus COVID-19

4.2.1 Uji Multikolinearitas

Uji multikolinearitas diperlukan untuk mengetahui apakah terdapat atribut-atribut yang memiliki hubungan kuat. Pada penelitian ini uji multikolinearitas dilakukan dengan melihat nilai koefisien korelasi *Pearson* antar atribut dan nilai *Variance Inflation Factor* (VIF). Lampiran 2 dan 3 menunjukkan nilai koefisien korelasi dan *p-value* antar kedua belas atribut. Koefisien korelasi antar atribut cukup beragam, terdapat beberapa atribut yang memiliki koefisien korelasi yang tinggi. Atribut jumlah rumah sakit umum dengan atribut jumlah dokter memiliki koefisien korelasi paling tinggi yaitu 0,94 dan *p-value* 0,00. Nilai ini menunjukkan bahwa peningkatan jumlah rumah sakit umum, maka jumlah dokter juga meningkat. Hal tersebut sesuai dengan kondisi sebenarnya, apabila jumlah rumah sakit banyak maka jumlah dokter yang ada juga banyak. Selain melihat nilai koefisien korelasi antar atribut, untuk mengetahui ada atau tidaknya hubungan yang linear antar variabel (multikolinearitas) dilihat berdasarkan nilai VIF pada masing-masing atribut.

Pada Lampiran 4 terdapat informasi mengenai nilai VIF setiap atribut. Agar tidak terdapat variabel yang mengindikasikan multikolinearitas pada data, maka variabel dengan nilai $VIF > 5$ harus dihilangkan. Proses penghilangan variabel yang mengindikasikan multikolinearitas dilakukan satu per satu. Pertama dihitung nilai VIF dari dua belas variabel sosial, ekonomi dan demografi yang akan digunakan. Terdapat beberapa variabel yang memiliki nilai

$VIF > 5$. Variabel jumlah dokter memiliki nilai VIF paling tinggi, variabel ini juga memiliki koefisien korelasi *Pearson* yang tinggi dengan beberapa atribut lainnya, salah satunya atribut jumlah rumah sakit umum, dengan koefisien korelasi *Pearson* 0,94. Hal ini menunjukkan bahwa variabel jumlah dokter, memiliki hubungan linear dengan variabel lainnya, oleh karena itu variabel ini tidak akan digunakan untuk proses analisis berikutnya. Pada langkah selanjutnya dilakukan kembali perhitungan VIF , dengan tidak mengikutsertakan variabel jumlah dokter. Hasil perhitungan VIF untuk kedua kalinya ini masih menunjukkan bahwa terdapat beberapa variabel yang memiliki nilai $VIF > 5$. Variabel total pengeluaran perkapita memiliki nilai VIF paling besar pada iterasi kedua, variabel ini juga memiliki korelasi nilai koefisien korelasi *Pearson* yang tinggi dengan beberapa variabel lainnya, salah satunya dengan variabel indeks pembangunan manusia. Oleh karena itu variabel total pengeluaran perkapita tidak akan digunakan pada analisis berikutnya. Proses yang sama dilakukan hingga tidak ada data yang memiliki nilai $VIF > 5$ atau korelasi *Pearson* $\geq 0,8$ atau korelasi *Pearson* $\geq -0,8$. Pada penelitian ini, setelah dilakukan lima iterasi didapatkan tujuh variabel yang tidak mengindikasikan adanya multikolinearitas yaitu, kepadatan penduduk, PDRB, jumlah populasi, tingkat pengangguran terbuka, persentase penduduk miskin, persentase penduduk lansia dan rasio jenis kelamin. Hanya variabel-variabel tersebut yang akan digunakan sebagai *input* untuk proses analisis berikutnya. Hasil perhitungan nilai VIF pada setiap iterasi terdapat pada Lampiran 4.

4.2.2 Pengelompokan Kabupaten/Kota di Jawa Timur Dengan Metode *Agglomerative Hierarchical Clustering*

Metode pengelompokan kabupaten/kota di Provinsi Jawa Timur yang digunakan pada penelitian ini yaitu metode berhirarki dengan pautan *Ward*. Pada langkah awal metode hirarki menganggap setiap objek sebagai *cluster*. Sehingga pada penelitian ini jumlah *cluster* awal adalah 38 *cluster*, yang merupakan jumlah kabupaten/kota di Provinsi Jawa Timur. Berdasarkan perhitungan jarak *Euclidean* yang terdapat pada Lampiran 5, nilai paling kecil adalah 0,776 yang merupakan jarak antar *cluster* objek 9 (Kabupaten Lamongan) dan objek 27 (Kabupaten Tuban). Kedua daerah dengan jarak terkecil ini digabung menjadi satu *cluster*. Berikutnya dilakukan perhitungan jarak antara *cluster* yang baru terbentuk dengan *cluster* lainnya sesuai dengan aturan pada pautan *Ward*. *Cluster* dengan nilai jarak minimum akan digabungkan. Proses ini diulangi lagi, hingga seluruh objek tergabung menjadi satu *cluster*. Informasi mengenai urutan pembentukan *cluster* yang disimpan pada matriks pautan ditampilkan secara lengkap pada Lampiran 6.

Secara eksploratif pautan *Ward* membentuk 2, 3, 4 dan 5 *cluster*, hal ini dapat diketahui berdasarkan dendrogram yang terdapat pada Lampiran 7. Solusi-solusi 2, 3, 4 dan 5 *cluster* memiliki jarak paling besar dibandingkan solusi lainnya. Akan tetapi, solusi yang akan digunakan untuk tahapan berikutnya hanya solusi 5 *cluster*. Solusi 5 *cluster* dipilih karena dibandingkan dengan solusi lainnya, solusi 5 *cluster* terlihat dapat membentuk *cluster* dengan jelas. Pada solusi 2 *cluster*, salah satu *cluster* terdiri dari 10 kabupaten/kota, sedangkan *cluster* lainnya terdiri dari 28 kabupaten/kota sisanya. Hasil akhir *clustering* menggunakan pautan *Ward* dengan lima *cluster* dapat dilihat pada Tabel 5.

Tabel 5 Hasil *clustering* metode hirarki dengan pautan *Ward*

Cluster	Jumlah	Anggota Cluster
Cluster 0	2	Kab. Sidoarjo, Kota Surabaya.
Cluster 1	13	Kab. Banyuwangi, Kab. Blitar, Kab. Bondowoso, Kab. Lumajang, Kab. Madiun, Kab. Magetan, Kab. Nganjuk, Kab. Ngawi, Kab. Pacitan, Kab. Ponorogo, Kab. Situbondo, Kab. Trenggalek, Kab. Tulungagung.
Cluster 2	8	Kota Batu, Kota Blitar, Kota Kediri, Kota Madiun, Kota Malang, Kota Mojokerto, Kota Pasuruan, Kota Probolinggo.
Cluster 3	5	Kab. Bangkalan, Kab. Pamekasan, Kab. Probolinggo, Kab. Sampang, Kab. Sumenep.
Cluster 4	10	Kab. Bojonegoro, Kab. Gresik, Kab. Jember, Kab. Jombang, Kab. Kediri, Kab. Lamongan, Kab. Malang, Kab. Mojokerto, Kab. Pasuruan, Kab. Tuban.

Hasil *clustering* dengan metode *Ward* dievaluasi dengan menghitung nilai koefisien *cophenetic*. Koefisien *cophenetic* mengukur korelasi antar elemen asli matriks jarak dengan elemen yang dihasilkan *cluster*. Hasil pengelompokan dikatakan baik jika nilainya mendekati angka 1. Nilai koefisien *cophenetic* yang dihasilkan dari proses *clustering* pada penelitian ini adalah 0,65. Nilai yang tidak terlalu tinggi, namun sudah dapat menunjukkan bahwa proses *clustering* yang dilakukan pada penelitian ini sudah cukup baik dan kabupaten/kota hasil *clustering* sudah berada pada *cluster* yang seharusnya.

4.2.3 Analisis Hasil Cluster

Hasil akhir dari proses *clustering* membagi kabupaten/kota di Jawa Timur ke dalam lima *cluster*. *Cluster* yang terbentuk dideskripsikan berdasarkan karakteristik sosial, ekonomi dan demografi. Nilai rata-rata setiap atribut pada masing-masing *cluster* dihitung untuk mengkategorikan *cluster* berdasarkan karakteristiknya. Nilai rata-rata setiap *cluster* pada tiap atribut dapat dilihat pada Tabel 6. Grafik pada Gambar 8 juga menampilkan visualisasi perbandingan nilai rata-rata dari data sosial, ekonomi dan demografi yang distandarisasi pada setiap *cluster*. Selain mengamati karakteristik sosial, ekonomi dan demografi pada masing-masing *cluster*, juga dilihat kondisi COVID-19 berdasarkan kategori jumlah kasus positif dan tingkat kematian COVID-19 di setiap wilayah anggota *cluster* per tanggal 30 Juni 2021.

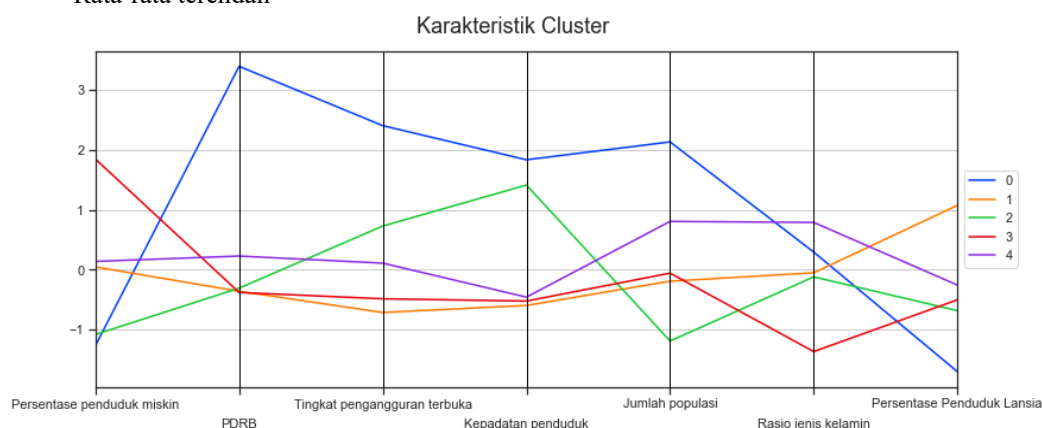
Tabel 6 Rata-rata atribut setiap *cluster*

Atribut	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Persentase penduduk miskin	5,305* (R)	11,225 (S)	6,139 (R)	19,346** (ST)	11,644 (S)
PDRB	263.120,8 75** (ST)	19.100,32 2 (S)	22.510,45 5 (S)	17.806,14 6* (S)	57.324,94 6 (S)

Tingkat pengangguran terbuka	10,380** (ST)	4,210* (R)	7,065 (T)	4,662 (R)	5,839 (S)
Kepadatan penduduk	5.741,5** (ST)	679,846* (R)	4.862,375 (T)	832,000 (R)	968,300 (S)
Jumlah populasi	2.478.558* (S)	942.975,2 (S)	283.508,1* (R)	1.031.420 (S)	1.602.474 (S)
Rasio jenis kelamin	99,870 (S)	99,187 (S)	99,051 (S)	96,582* (R)	100,85** (T)
Persentase Penduduk Lansia	8,675* (SR)	17,111** (ST)	11,771 (S)	12,32 (S)	13,068 (S)

** Rata-rata tertinggi

* Rata-rata terendah



Gambar 8 Koordinat paralel rata-rata atribut setiap *cluster*

Cluster Nol

Cluster nol adalah *cluster* dengan jumlah anggota paling sedikit, hanya terdiri dari dua daerah yaitu Kabupaten Sidoarjo dan Kota Surabaya. Kedua wilayah ini secara geografis berbatasan secara langsung. Berdasarkan Gambar 8, diketahui bahwa *cluster nol* memiliki karakteristik unik, yaitu sebagian besar nilai rata-rata atributnya selalu paling tinggi dibandingkan *cluster* lainnya. *Cluster* ini memiliki nilai PDRB, tingkat pengangguran terbuka, kepadatan penduduk, dan jumlah populasi paling tinggi. Kedua daerah pada *cluster nol* merupakan pusat aktivitas perekonomian di Jawa Timur (Bank Dunia 2012). Daerah ini memiliki corak industri yang cukup dominan yang ditandai dengan berdirinya perusahaan-perusahaan besar seperti perusahaan rokok Sampoerna, Unilever, Bogasari, industri emas UBS, dsb. Selain itu pada daerah ini sektor perdagangan, hotel dan restoran juga berkembang dengan pesat

Namun, nilai PDRB yang tinggi tidak diikuti dengan rendahnya nilai tingkat pengangguran. *Cluster nol* memiliki tingkat pengangguran paling tinggi dibandingkan *cluster* lainnya, kondisi ini seringkali terjadi di daerah industri. Tingginya tingkat pengangguran di daerah dengan PDRB tinggi disebabkan karena terjadinya peningkatan jumlah produksi, namun tidak diikuti dengan

penyerapan tenaga kerja yang maksimal. Berbagai faktor dapat menyebabkan hal tersebut terjadi, diantaranya karena adanya perbedaan tingkat keahlian dan tingkat pendidikan yang dibutuhkan industri dengan tenaga kerja yang tersedia. Dengan kondisi perekonomian tersebut, *cluster* nol memiliki persentase penduduk miskin yang rendah, yang menunjukkan bahwa sebagian besar masyarakat pada wilayah ini sudah berada pada kondisi ekonomi yang cukup baik. Secara demografis, *cluster* nol memiliki jumlah populasi dan kepadatan penduduk paling tinggi dibandingkan wilayah lainnya. Rasio jenis kelamin 99,87 memiliki makna bahwa ada sekitar 99 penduduk laki-laki setiap 100 penduduk perempuan. *Cluster* nol memiliki persentase penduduk lansia yang sangat rendah, yang berarti bahwa sebagian besar penduduk pada wilayah ini berada pada usia produktif (15-64 tahun).

Jika dilihat berdasarkan kondisi COVID-19 pada Tabel 7, kedua wilayah *cluster* nol memiliki jumlah kasus positif COVID-19 yang sangat tinggi. Hal ini berkaitan dengan tingginya terjadinya pergerakan manusia, karena karakteristik ekonomi dan demografi di wilayah ini. Tingginya nilai PDRB memungkinkan terjadinya aktivitas perekonomian yang tinggi. Selain itu kepadatan penduduk yang padat memungkinkan terjadinya kontak atau interaksi secara langsung juga tinggi. Namun, tingkat kematian karena COVID-19 di *cluster* nol ini rendah. Hal ini mungkin disebabkan karena wilayah ini memiliki persentase penduduk lansia yang sangat rendah. Diketahui bahwa penduduk lansia memiliki tingkat kekebalan tubuh lebih lemah dibandingkan penduduk usia produktif. Hal ini berisiko infeksi COVID-19 menimbulkan gangguan yang parah bahkan kematian pada lansia. Namun, karena persentase lansianya yang sangat rendah, kematian karena COVID-19 di *cluster* nol juga rendah.

Tabel 7 Kondisi COVID-19 di wilayah *cluster* nol

Kabupaten/Kota	Jumlah kasus positif COVID-19	Tingkat kematian karena COVID-19
Kabupaten Sidoarjo	Sangat tinggi	Rendah
Kota Surabaya	Sangat tinggi	Rendah

Cluster satu

Cluster satu adalah *cluster* dengan jumlah anggota paling banyak yang terdiri dari tiga belas kabupaten yaitu, Banyuwangi, Blitar, Bondowoso, Lumajang, Madiun, Magetan, Ngawi, Pacitan, Ponorogo, Situbondo, Trenggalek, dan Tulungagung. Sebagian besar daerah pada *cluster* nol ini merupakan daerah-daerah paling luar dari Provinsi Jawa Timur. Karakteristik unik dari daerah ini adalah, persentase penduduk lansianya paling tinggi, tingkat pengangguran terbuka serta kepadatan penduduknya paling rendah. Berdasarkan data BPS, lapangan pekerjaan utama pada wilayah ini ada pada sektor pertanian dan perdagangan, sebanyak sembilan kabupaten sebagian besar masyarakatnya bekerja pada bidang pertanian dan pada empat kabupaten lainnya sektor perdagangan menjadi lapangan pekerjaan utama bagi masyarakat. Karena sektor pertanian dan perdagangan paling dominan pada daerah ini, penyerapan tenaga kerja dapat lebih maksimal, sehingga tingkat pengangguran terbuka pada *cluster*

ini rendah. Berdasarkan karakteristik perekonomian, wilayah ini memiliki nilai PDRB yang rendah dengan persentase penduduk miskin cukup tinggi.

Berdasarkan Tabel 8, sebelas kabupaten pada *cluster* satu ada pada kategori sedang untuk jumlah terkonfirmasi positif COVID-19, satu kabupaten pada kategori rendah dan satu kabupaten pada kategori tinggi. Kabupaten Banyuwangi merupakan kabupaten dengan kategori kasus COVID-19 tinggi. Jika dilihat pada data sosial, ekonomi dan demografinya, Kabupaten Banyuwangi merupakan daerah dengan PDRB dan jumlah populasi paling tinggi pada *cluster* ini. Hal ini mungkin berkaitan dengan tingginya jumlah kasus positif COVID-19 di sana. Sedangkan Kabupaten Ngawi merupakan daerah dengan jumlah kasus positif COVID-19 kategori rendah. Karakteristik sosial, ekonomi dan demografi dari Kabupaten Ngawi ini yaitu, merupakan daerah dengan persentase penduduk miskin paling tinggi dan kepadatan penduduk paling rendah di wilayah *cluster* satu ini. Pada tingkat kematian karena COVID-19 satu kabupaten kategori sangat tinggi, lima kabupaten kategori tinggi, empat kabupaten kategori sedang, dua kabupaten lainnya kategori sangat rendah.

Tabel 8 Kondisi COVID-19 di wilayah *cluster* satu

Kabupaten/Kota	Jumlah kasus positif COVID-19	Tingkat kematian karena COVID-19
Kabupaten Banyuwangi	Tinggi	Tinggi
Kabupaten Blitar	Sedang	Sangat tinggi
Kabupaten Bondowoso	Sedang	Sedang
Kabupaten Lumajang	Sedang	Sedang
Kabupaten Madiun	Sedang	Rendah
Kabupaten Magetan	Sedang	Sedang
Kabupaten Nganjuk	Sedang	Tinggi
Kabupaten Ngawi	Rendah	Tinggi
Kabupaten Pacitan	Sedang	Sangat rendah
Kabupaten Ponorogo	Sedang	Tinggi
Kabupaten Situbondo	Sedang	Sedang
Kabupaten Trenggalek	Sedang	Tinggi
Kabupaten Tulungagung	Sedang	Sangat rendah

Cluster dua

Cluster dua terdiri dari seluruh wilayah kota di Jawa Timur kecuali kota Surabaya. Anggota *cluster* ini yaitu Kota Batu, Blitar, Kediri, Madiun, Malang, Mojokerto, Pasuruan dan Probolinggo. Wilayah ini memiliki tingkat pengangguran dan kepadatan penduduk yang tinggi, namun jumlah populasinya paling rendah. Hal ini disebabkan karena luas wilayah pada anggota *cluster* dua ini cukup sempit dibandingkan wilayah *cluster* lainnya sehingga kepadatan penduduk tinggi meskipun jumlah penduduknya tidak cukup banyak. Nilai rata-rata PDRB di wilayah ini ada pada kategori sedang. Pertumbuhan ekonomi yang

belum terlalu pesat di wilayah ini disebabkan karena adanya daerah pada *cluster* ini merupakan daerah otonom baru sehingga masih bergantung pada pemerintah pusat untuk menunjang adanya pertumbuhan ekonomi (Putri 2017). Sektor perdagangan, jasa, dan pariwisata adalah sektor yang banyak berkontribusi terhadap nilai PDRB pada wilayah ini. Berdasarkan kondisi sosial dan demografi, persentase penduduk miskin pada kategori rendah, hal ini menunjukkan bahwa sebagian masyarakat pada daerah ini berada pada kondisi perekonomian cukup baik. Selain itu, persentase penduduk lansia dan rasio jenis kelamin berada pada kategori sedang. Rasio kelamin 99,05 menunjukkan bahwa jumlah penduduk perempuan lebih banyak dibandingkan penduduk laki-laki.

Berdasarkan kondisi COVID-19 pada Tabel 9, sebagian besar anggota *cluster* dua berada pada kategori rendah dan sedang, hanya satu kota yang ada pada kategori tinggi, yaitu Kota Malang. Jika diamati lebih lanjut, Kota Malang memiliki karakteristik tingkat pengangguran terbuka, kepadatan penduduk, dan jumlah populasi paling tinggi dibandingkan daerah lainnya pada *cluster* dua. Berdasarkan tingkat kematian karena COVID-19, kategori setiap daerah sangat bervariasi mulai dari rendah hingga tinggi.

Tabel 9 Kondisi COVID-19 di wilayah *cluster* dua

Kabupaten/Kota	Jumlah kasus positif COVID-19	Tingkat kematian karena COVID-19
Kota Batu	Rendah	Tinggi
Kota Blitar	Sedang	Rendah
Kota Kediri	Rendah	Tinggi
Kota Madiun	Sedang	Sedang
Kota Malang	Tinggi	Tinggi
Kota Mojokerto	Sedang	Sedang
Kota Pasuruan	Rendah	Tinggi
Kota Probolinggo	Rendah	Sedang

Cluster tiga

Cluster tiga terdiri dari Kabupaten Bangkalan, Pamekasan, Probolinggo, Sampang dan Sumenep. Hampir seluruh daerah di *cluster* tiga ini terletak di Pulau Madura dan kepulauan sekitarnya, kecuali Kabupaten Probolinggo. Berdasarkan Tabel 7, *cluster* tiga memiliki rata-rata persentase penduduk miskin paling tinggi dan nilai PDRB yang rendah. Hal ini menunjukkan bahwa kesejahteraan masyarakat dan kondisi perekonomian belum terlalu baik pada wilayah *cluster* tiga. Posisinya yang terpisah dari pulau Jawa menyebabkan adanya keterbatasan akses terhadap pelayanan publik atau kurang berkembangnya kegiatan perekonomian di wilayah ini (Bank Dunia 2012). Sektor pertanian, kehutanan dan perikanan merupakan basis utama perekonomian pada daerah ini. Namun sektor ini tidak menyumbang nilai yang terlalu tinggi pada PDRB, tidak seperti sektor perindustrian atau perdagangan. Namun, pada wilayah ini tingkat pengangguran terbuka sangat rendah. Hal tersebut berkaitan dengan sektor perekonomian utama pada *cluster* tiga adalah pertanian dan perikanan, sehingga penyerapan tenaga kerja dapat lebih maksimal. Secara demografis *cluster* tiga memiliki jumlah populasi dan persentase

penduduk lansia yang termasuk pada kategori sedang. Sedangkan tingkat kepadatan penduduk pada wilayah ini berada pada kategori rendah. Untuk rataan rasio jenis kelamin, *cluster* tiga memiliki nilai paling rendah dibandingkan *cluster* lainnya yaitu 96, 582. Nilai ini memiliki makna bahwa terdapat 96 penduduk laki-laki per 100 penduduk perempuan. Hal tersebut menunjukkan bahwa jumlah penduduk perempuan lebih banyak dibandingkan penduduk laki-laki pada *cluster* ini.

Berdasarkan Tabel 10 untuk kondisi COVID-19, seluruh wilayah yang berada pada Pulau Madura dan kepulauan sekitarnya kecuali Kabupaten Bangkalan berada pada kategori rendah, sedangkan Kabupaten Probolinggo dan Kabupaten Bangkalan menjadi wilayah pada *cluster* ini yang termasuk pada kategori sedang. Jika dilihat pada data sosial, ekonomi dan demografi Kabupaten Probolinggo memiliki jumlah populasi paling tinggi dibandingkan daerah lainnya di *cluster* ini. Sedangkan Kabupaten Bangkalan adalah daerah dengan tingkat pengangguran paling tinggi di wilayah *cluster* ini. Berdasarkan tingkat kematian karena COVID-19 tiga kabupaten ada pada kategori rendah, satu kabupaten kategori sedang dan satu kabupaten kategori tinggi. Kabupaten dengan tingkat kematian karena COVID-19 kategori tinggi pada *cluster* ini yaitu Kabupaten Bangkalan.

Tabel 10 Kondisi COVID-19 di wilayah *cluster* tiga

Kabupaten/Kota	Jumlah kasus positif COVID-19	Tingkat kematian karena COVID-19
Kabupaten Bangkalan	Sedang	Tinggi
Kabupaten Pamekasan	Rendah	Sedang
Kabupaten Probolinggo	Sedang	Rendah
Kabupaten Sampang	Rendah	Rendah
Kabupaten Sumenep	Rendah	Rendah

Cluster empat

Cluster empat terdiri dari sepuluh kabupaten yaitu, Bojonegoro, Gresik, Jember, Jombang, Kediri, Lamongan, Malang, Mojokerto, Pasuruan, dan Tuban. Atribut persentase penduduk miskin, tingkat pengangguran terbuka, PDRB, kepadatan penduduk, jumlah populasi dan persentase penduduk lansia termasuk kategori sedang. Hanya atribut rasio jenis kelamin yang termasuk kategori tinggi. Pada *cluster* empat nilai rata-rata rasio jenis kelaminnya adalah 100,85. Artinya terdapat 100 penduduk laki-laki per 100 orang penduduk perempuan. Hal ini menunjukkan bahwa *cluster* empat memiliki perbandingan penduduk laki-laki dan perempuan paling seimbang dibandingkan *cluster* lainnya. Pada sektor perekonomian, *cluster* empat memiliki lapangan pekerjaan utama pada bidang pertanian dan perdagangan.

Berdasarkan Tabel 11 untuk kondisi COVID-19, hampir seluruh daerah pada *cluster* empat termasuk pada kategori sedang, hanya Kabupaten Jember yang termasuk pada kategori tinggi. Namun, jika dilihat kembali pada data sosial, ekonomi dan demografi tidak ada karakteristik yang terlalu berbeda dari Kabupaten Jember dibandingkan dengan kabupaten lainnya pada *cluster* empat ini. Kabupaten Jember memiliki nilai rasio jenis kelamin paling kecil

dibandingkan daerah lainnya, yaitu 99,47. Nilai ini menunjukkan bahwa, pada Kabupaten Jember jumlah penduduk perempuan lebih banyak dibandingkan penduduk laki-laki. Berdasarkan tingkat kematian karena COVID-19, wilayah di *cluster* empat terbagi menjadi berbagai kategori yaitu rendah, sedang, dan tinggi.

Tabel 11 Kondisi COVID-19 di wilayah *cluster* empat

Kabupaten/Kota	Jumlah kasus positif COVID-19	Tingkat kematian karena COVID-19
Kabupaten Bojonegoro	Sedang	Rendah
Kabupaten Gresik	Sedang	Rendah
Kabupaten Jember	Tinggi	Sedang
Kabupaten Jombang	Sedang	Tinggi
Kabupaten Kediri	Sedang	Tinggi
Kabupaten Lamongan	Sedang	Rendah
Kabupaten Malang	Sedang	Sedang
Kabupaten Mojokerto	Sedang	Sangat rendah
Kabupaten Pasuruan	Sedang	Sedang
Kabupaten Tuban	Sedang	Tinggi

4.2.4 Analisis Regresi Multivariat

Analisis regresi dilakukan untuk mengetahui pengaruh variabel bebas (atribut sosial, ekonomi dan demografi) terhadap variabel terikat (jumlah terkonfirmasi positif COVID-19).

Tabel 12 Hasil analisis linear berganda

Atribut	Koefisien regresi	p-value
PDRB	0,6668	0,000
Persentase penduduk lansia	0,3258	0,001
Jumlah populasi	0,2779	0,015
Kepadatan penduduk	0,1977	0,142
Tingkat pengangguran terbuka	0,1912	0,060
Persentase penduduk miskin	-0,0304	0,749
Rasio jenis kelamin	-0,0401	0,599

*berpengaruh nyata pada taraf 1%

Berdasarkan analisis regresi seperti yang tersaji pada Tabel 12, dapat dilihat bahwa hanya atribut PDRB yang memiliki pengaruh nyata terhadap jumlah terkonfirmasi positif COVID-19. Atribut PDRB memiliki nilai koefisien regresi paling tinggi dibandingkan atribut lainnya, hal ini menunjukkan bahwa dari tujuh variabel bebas yang digunakan pada analisis regresi ini, PDRB merupakan atribut yang paling berpengaruh dalam menentukan jumlah terkonfirmasi positif COVID-19 di Provinsi Jawa Timur. Jika diamati pada hasil analisis *cluster* yang telah dilakukan sebelumnya, bahwa kasus positif COVID-19 yang tinggi terdapat pada daerah dengan nilai Produk Domestik Regional

Bruto (PDRB) juga tinggi. Nilai PDRB yang tinggi menjadi suatu ciri daerah dengan aktivitas perekonomian yang sangat aktif, di mana penduduknya selalu bergerak. Sehingga kemungkinan terinfeksi virus lebih signifikan apabila tidak dilakukan pengendalian. Selain itu, daerah dengan kekuatan ekonomi yang baik, juga dapat melakukan tes lebih banyak dan menjangkau masyarakat lebih luas, sehingga kasus positif dapat terdeteksi lebih banyak.

Koefisien determinasi dapat menjelaskan variasi pengaruh variabel bebas yaitu PDRB, jumlah populasi, persentase penduduk lansia, kepadatan penduduk, tingkat pengangguran terbuka, rasio jenis kelamin, dan persentase penduduk miskin terhadap variabel terikat yaitu jumlah terkonfirmasi positif COVID-19. Nilai koefisien determinasi diukur oleh nilai *adjusted R-squared* untuk variabel bebas lebih dari satu seperti pada Tabel 13.

Tabel 13 Koefisien determinasi

<i>R-squared</i>	0,890
<i>Adjusted R-squared</i>	0,864

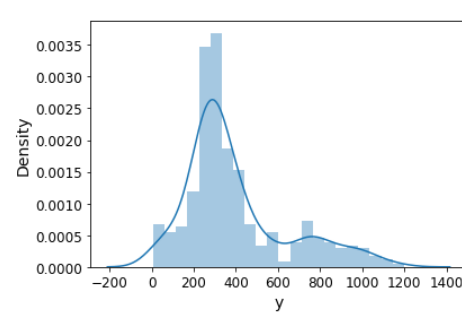
Nilai *Adjusted R-squared* 0,864 menjelaskan bahwa 86,4% proporsi pengaruh dari variabel bebas yang ada pada model terhadap variabel terikatnya, sedangkan 13,6% lainnya dipengaruhi oleh variabel lain yang tidak terdapat pada model regresi linear berganda ini. Hal ini disebabkan karena variabel bebas yang digunakan pada penelitian ini hanya beberapa faktor sosial, ekonomi dan demografi, sedangkan ada banyak variabel lainnya yang dapat mempengaruhi jumlah terkonfirmasi positif COVID-19 di suatu daerah.

4.3 Prediksi Penambahan Jumlah terkonfirmasi positif COVID-19 di Jawa Timur

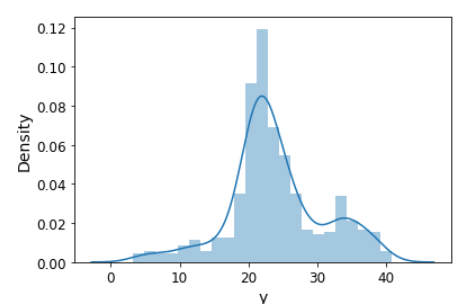
Prediksi penambahan jumlah terkonfirmasi positif COVID-19 di Jawa Timur dilakukan untuk mendapatkan gambaran mengenai kemungkinan kondisi COVID-19 yang akan terjadi untuk periode waktu tertentu, pada penelitian ini periode waktu yang akan diprediksi adalah tiga puluh hari.

4.3.1 Pra-proses data

Pada tahapan pra-proses, data awal yang merupakan data positif COVID-19 per kabupaten/kota dilakukan agregasi agar diperoleh data harian positif COVID-19 untuk satu Provinsi Jawa Timur. Berikutnya dilakukan proses transformasi pada data. Tujuan dari proses transformasi ini adalah agar nilai ragam pada data lebih kecil dan membuat data mendekati distribusi normal. Pada analisis data deret waktu, data dengan ragam tinggi, menyebabkan model akan menangkap lebih banyak error, error pada titik masa lalu juga akan kembali terlihat di masa akan datang. Oleh karena itu, pada penelitian ini teknik transformasi yang akan digunakan adalah transformasi Box-Cox. Transformasi Box Cox mengevaluasi sekumpulan koefisien lambda (λ) dan memilih nilai yang mencapai pendekatan normalitas terbaik.



Gambar 9 Plot distribusi penambahan positif harian COVID-19 sebelum ditransformasi



Gambar 10 Plot distribusi data setelah ditransformasi

Gambar 9 merupakan plot distribusi data sebelum ditransformasi data ini memiliki nilai simpangan baku cukup besar yaitu 244,86. Sedangkan Gambar 10 menunjukkan distribusi data setelah dilakukan transformasi Box-Cox, terlihat data ini sudah mendekati bentuk distribusi normal dengan nilai simpangan baku yang lebih kecil yaitu 6,86.

4.3.2 Pemodelan Prediksi Menggunakan *FBProphet*

Pada tahapan ini dilakukan dua kali pemodelan, pemodelan pertama menggunakan parameter *default* dari *FBProphet*. Sedangkan pada pemodelan yang kedua dilakukan proses *hyperparameter* tuning dengan *grid search* untuk menemukan parameter terbaik. Pada Tabel 14 terdapat nilai terbaik dari proses *hyperparameter* tuning.

Tabel 14 Daftar parameter yang dilakukan *tuning*

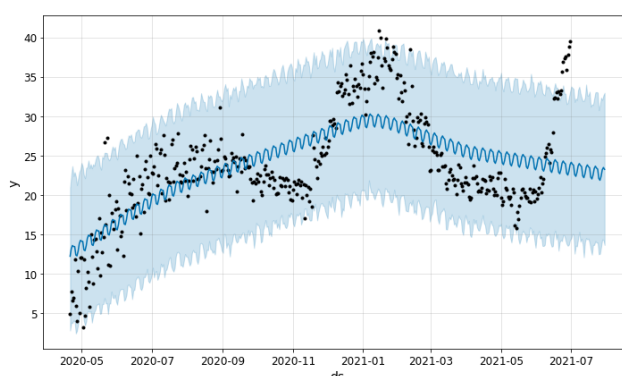
<i>Hyperparameter</i>	Nilai <i>tuning</i>	Nilai terbaik
changepoint_prior_scale	[0.001, 0.01, 0.05, 0.1, 0.25, 0.5]	0.5
changepoint_range	[0.8, 0.9, 0.95]	0.95
seasonality_prior_scale	[0.01, 5, 10]	10

Changepoint_prior_scale dan changepoint_range adalah dua parameter dari komponen *trend* pada *FBProphet*. Changepoint_prior_scale menentukan fleksibilitas dari garis *trend* yang akan dibentuk, semakin besar nilainya maka *trend* yang dibentuk semakin fleksibel. Pada penelitian ini, nilai

changeoint_prior_scale terbaik untuk model adalah 0,5. Kemudian, parameter changeoint_range menentukan proporsi dari data yang memungkinkan *trend* mengalami perubahan. Nilai changeoint_range terbaik untuk model prediksi ini adalah 0,95 yang berarti, *trend* dapat berubah pada 95% data awal. Seasonality_prior_scale merupakan parameter dari komponen *seasonality* pada *FBProphet*. Parameter ini menentukan fleksibilitas dari efek musiman pada data. Nilai seasonality_prior_scale terbaik untuk model pada penelitian ini adalah 10.

Pada proses pemodelan juga ditambahkan komponen *holiday*. Komponen *holiday* ditambahkan karena berdasarkan data diketahui bahwa jumlah terkonfirmasi positif COVID-19 meningkat dengan tajam pada hari atau peristiwa tertentu. Rizal (2020) menyatakan bahwa masa libur panjang sering mendorong masyarakat untuk bepergian, padahal pergerakan dan kerumunan bisa berdampak pada meningkatnya jumlah terkonfirmasi positif COVID-19. Menurut data Satgas Penanganan COVID-19 pada setiap periode liburan selalu terjadi *trend* kenaikan jumlah kasus.

Berikutnya, untuk melakukan validasi model digunakan teknik *expanding window cross validation*. Pada *cross validation* ini jumlah data *train* awal adalah 280 hari, data *test* atau *horizon* 30 hari, dan penambahan data *train* pada setiap peningkatan *fold* adalah 30 hari. Plot hasil prediksi menggunakan kedua model ditunjukkan pada Gambar 11 dan 12.



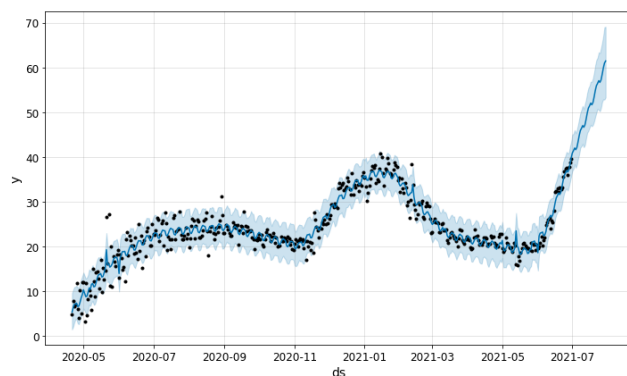
Gambar 11 Plot prediksi menggunakan model *default FBProphet*

Gambar 11 merupakan plot hasil prediksi menggunakan model *default FBProphet*. Terlihat bahwa model dengan parameter *default FBProphet* tidak mampu mengikuti pola pergerakan dari data aktual. Dari April 2020 - Januari 2021, model prediksi ini menganggap bahwa hanya terjadi kenaikan pada data. Sedangkan pada data aktual, terlihat bahwa terjadi kenaikan juga penurunan kasus positif COVID-19. Secara *default FBProphet* mengatur nilai parameter changeoint_prior_scale menjadi 0,05 dan changeoint_range menjadi 0,8. Nilai tersebut merupakan nilai paling kecil untuk kedua parameter. Changeoint_prior_scale 0,05 memiliki arti bahwa *trend* perubahan pada data kurang fleksibel. Sedangkan changeoint_range 0,8 memiliki arti bahwa perubahan *trend* hanya dibolehkan terjadi pada 80% data awal. Jika diamati pada data yang digunakan, berarti perubahan *trend* hanya boleh terjadi sampai dengan tanggal 5 April 2021. Sedangkan pada *trend* keseluruhan data, perubahan tidak hanya terjadi hingga periode tersebut, setelah periode April 2021 juga terjadi

perubahan *trend* yang signifikan. Oleh karena itu, model yang digunakan juga perlu mempelajari perubahan ini. Hal-hal tersebut menyebabkan model tidak dapat menangkap perubahan *trend* pada data dengan baik.

Pada komponen *seasonality*, model *default FBProphet* mengatur nilai parameter *seasonality_prior_scale* menjadi 10. Nilai ini merupakan nilai terbesar untuk parameter *seasonality_prior_scale*, yang memungkinkan model untuk lebih menyesuaikan fluktuasi dari efek musiman pada data. Jika diamati lebih lanjut, pada Gambar 11 terlihat bahwa pada satu bulan terdapat setidaknya 8 bukit atau lembah, yang menunjukkan efek *seasonality* pada data.

Pada model *default FBProphet* belum memperhitungkan pengaruh dari komponen *holiday*. Hal ini menyebabkan model tidak dapat memprediksi dengan baik pada data-data yang nilainya sangat berbeda dengan data pada hari biasanya. Contohnya di bulan Mei 2020, pada Gambar 11 terlihat ada data yang nilainya jauh lebih tinggi dibandingkan data pada hari lainnya. Namun, model prediksi *default FBProphet* tidak dapat menjangkau nilai tersebut. Pada bulan Mei 2020 ini terdapat beberapa hari libur yang berkemungkinan memiliki pengaruh terhadap data positif COVID-19 yaitu, Hari Buruh, Hari Waisak, Hari Kenaikan Yesus Kristus, dan Hari Raya Idul Fitri.



Gambar 12 Plot prediksi menggunakan model *FBProphet hyperparameter tuning*

Gambar 12 merupakan plot hasil prediksi menggunakan model *FBProphet* yang telah dilakukan *hyperparameter tuning*. Terlihat bahwa hasil prediksi sudah dapat mengikuti pergerakan dari data aktual dengan baik. Tabel 15 sebelumnya menunjukkan nilai pada parameter yang dilakukan *tuning*. Jika dibandingkan dengan plot hasil prediksi dari model *default FBProphet* yang ditunjukkan Gambar 11. Prediksi menggunakan model yang sudah dilakukan *hyperparameter tuning* ini terlihat lebih baik dalam memprediksi penambahan jumlah positif COVID-19. Model ini mampu mengikuti perubahan *trend* pada data dengan baik. *Trend* pada data hasil prediksi terlihat fluktuatif, sesuai dengan kondisi *trend* pada data aktual. Pada model prediksi ini sudah memperhitungkan pengaruh dari komponen *holiday*. Hal ini juga mempengaruhi hasil prediksi, model dapat memperkirakan perubahan pada data dengan lebih baik. Contohnya pada Mei 2020, yang terdapat beberapa hari libur pada periode tersebut, pada hasil prediksi ada titik yang nilainya sangat naik dibandingkan hari lainnya. Meskipun belum mencapai nilai pada data aktual, namun hasil prediksi ini sudah baik karena dapat memperkirakan adanya kemungkinan kenaikan atau penurunan data yang sangat berbeda dibandingkan

dengan data pada hari biasanya. Selain melihat pada plot hasil prediksi kedua model, perbandingan performa dari model *default FBprophet* dan *FBProphet* yang dilakukan *hyperparameter tuning* juga dilihat berdasarkan nilai *error* yang dihasilkan. Pada Tabel 15 terdapat perbandingan nilai RMSLE untuk kedua model prediksi.

Tabel 15 Perbandingan model prediksi

Model	RMSLE
<i>Default FBProphet</i>	0,38
<i>FBProphet dengan hyperparameter tuning</i>	0,25

Nilai RMSLE menunjukkan jarak rata-rata antara nilai pada model prediksi dan nilai aktual pada *dataset*. Berdasarkan Tabel 16, *error* pada model menjadi lebih kecil setelah dilakukan *hyperparameter tuning*. Sehingga model ini akan digunakan untuk melakukan prediksi penambahan jumlah terkonfirmasi positif COVID-19 di Jawa Timur untuk periode tiga puluh hari berikutnya.

4.3.3 Prediksi Positif COVID-19 Untuk Periode Berikutnya

Gambar 12 merupakan hasil plot prediksi COVID-19 di Jawa Timur pada data yang dilakukan transformasi Box-Cox. Hasil prediksi dari model *FBProphet* merupakan penjumlahan dari komponen-komponennya. Untuk membuktikan hal ini maka dilakukan perhitungan dengan menjumlahkan seluruh nilai prediksi (*yhat*) yaitu nilai prediksi dari tanggal 21 April 2020 sampai dengan 30 Juli 2021. Kemudian dilakukan juga perhitungan dengan menjumlahkan seluruh komponen *FBProphet* yang terdapat pada data (*trend*, *weekly*, *holiday*). Hasil penjumlahan ini akan dibandingkan, apakah nilainya sama atau tidak. Tabel 16 merupakan cuplikan hasil prediksi menggunakan *FBProphet*.

Tabel 16 Hasil prediksi dan komponen *FBProphet*

date	<i>trend</i>	<i>weekly</i>	<i>holiday</i>	<i>yhat</i>
21 April 2020	5,75656	-0,35851	0	5,39805
22 April 2020	6,07054	0,32600	0	6,39654
⋮	⋮	⋮	⋮	⋮
01 Mei 2020	8,89632	0,49925	0,97204	10,36761
02 Mei 2020	9,21029	0,38943	0	9,59972
⋮	⋮	⋮	⋮	⋮
total	11936,93	1,19	2,15	11940,27

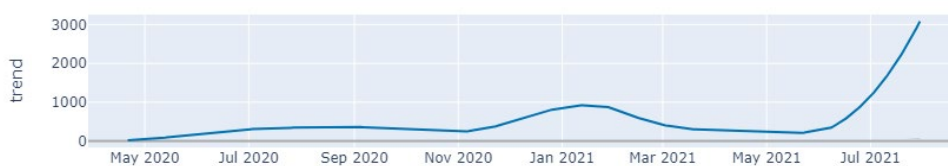
Berdasarkan Tabel 16 terlihat bahwa nilai prediksi (*yhat*) pada *FBProphet* merupakan penjumlahan dari komponen *trend*, *weekly* dan *holiday*. Nilai ini masih merupakan nilai pada data yang dilakukan transformasi Box-Cox. Untuk memperoleh nilai aktual maka perlu dilakukan invers pada data. Gambar 13 merupakan plot hasil prediksi positif COVID-19 menggunakan *FBProphet* pada data yang sudah dilakukan invers.



Gambar 13 Plot prediksi penambahan positif harian COVID-19 hingga 30 Juli 2021

Berdasarkan Gambar 13 diperkirakan bahwa untuk periode tiga puluh hari berikutnya penambahan jumlah terkonfirmasi positif COVID-19 di Provinsi Jawa Timur masih terus naik. Jika tidak dilakukan tindakan pengendalian penyebaran COVID-19 yang lebih ketat, diperkirakan pada tanggal 30 Juli 2021 terjadi penambahan jumlah positif COVID-19 di Provinsi Jawa Timur sebanyak 3.146 kasus baru.

4.3.4 Analisis Komponen *FBProphet*



Gambar 14 Plot komponen *trend* pada *FBProphet*

Ada tiga komponen yang mempengaruhi nilai prediksi pada *FBProphet*, yaitu *trend*, *holiday*, dan *weekly (seasonality)*. Gambar 14 merupakan plot untuk komponen *trend*. Pada komponen *trend*, pertambahan jumlah terkonfirmasi positif COVID-19 cukup fluktuatif dari April 2020 sampai dengan Juli 2021. Terlihat bahwa terjadi kenaikan kasus positif dari April – Juli 2020, kemudian terjadi penurunan dari September – November 2020. Namun, *trend* kasus positif COVID-19 di Jawa Timur kembali naik di bulan Oktober 2020 hingga Februari 2021. Kemudian kembali turun di tiga bulan berikutnya. Diperkirakan pada Juni-Juli 2021 *trend* penambahan terus mengalami kenaikan yang signifikan.



Gambar 15 Plot komponen *holiday* pada *FBProphet*

Pada Gambar 15 terdapat plot untuk komponen *holidays*, ada beberapa hari libur yang memiliki pengaruh sangat tinggi terhadap kenaikan jumlah terkonfirmasi positif COVID-19 di Jawa Timur diantaranya yaitu hari besar keagamaan (Hari Raya Idul Fitri, Hari Kenaikan Yesus Kristus dan Tahun Baru Imlek), serta Hari Buruh. Adanya pengaruh dari hari libur panjang terhadap penambahan jumlah terkonfirmasi positif COVID-19 berkaitan dengan terjadinya peningkatan aktivitas masyarakat pada hari tersebut, sehingga kemungkinan tertular COVID-19 juga dapat meningkat. Namun, ada beberapa hari libur lainnya yang justru memiliki pengaruh kecil terhadap penambahan jumlah positif COVID-19 di Jawa Timur. Contohnya adalah pada Hari Raya Waisak dan Isra Mi'raj, tahun baru Islam, Maulid Nabi Muhammad SAW. Hari libur tersebut bukan termasuk hari libur panjang dan sebagian besar jatuh di tengah hari kerja yaitu Hari Rabu. Karena bukan termasuk hari libur panjang, kemungkinan aktivitas pergerakan masyarakat pada hari tersebut tidak terlalu meningkat dibandingkan dengan hari lainnya. Sehingga penambahan jumlah positif COVID-19 juga tidak terlalu signifikan.



Gambar 16 Plot komponen *seasonality* pada *FBProphet*

Pada komponen *seasonality*, berdasarkan data yang tersedia hanya *weekly effect* yang dapat diamati. Hari Kamis memiliki efek paling tinggi terhadap penambahan jumlah terkonfirmasi positif COVID-19 dibandingkan hari lainnya. Sedangkan akhir pekan memiliki pengaruh paling rendah terhadap penambahan jumlah terkonfirmasi positif. Menurut Hale (2020), rendahnya jumlah terkonfirmasi positif COVID-19 di akhir pekan disebabkan oleh beberapa faktor diantaranya adalah jumlah tes yang dilakukan di akhir pekan lebih sedikit, laboratorium memproses lebih sedikit tes pada akhir pekan, adanya keterlambatan pelaporan oleh departemen kesehatan, serta pada akhir pekan biasanya masyarakat sering melakukan pertemuan dengan kerabat yang menyebabkan lebih banyak gejala muncul di hari kerja.

V SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil yang diperoleh dari analisis deskriptif, distribusi penyebaran jumlah terkonfirmasi positif COVID-19 di Jawa Timur sangat beragam. Jumlah kasus yang berbeda-beda di setiap daerah bisa disebabkan karena berbagai faktor, termasuk faktor sosial, ekonomi dan demografi di setiap daerah. Jumlah terkonfirmasi positif COVID-19 terbanyak ditemukan di daerah pusat perekonomian dengan tingkat kepadatan penduduk yang tinggi seperti Kota Surabaya, Kabupaten Sidoarjo serta wilayah yang terhubung dengan daerah ini. Sedangkan kasus terendah ditemukan di wilayah yang basis utama

perekonomiannya adalah pertanian, perkebunan dan perikanan dengan jumlah dan tingkat kepadatan penduduk yang rendah. Berdasarkan hasil analisis regresi dengan nilai *adjusted r-squared* 0,864 diketahui bahwa atribut PDRB merupakan atribut yang berpengaruh signifikan terhadap jumlah kasus COVID-19 di Jawa Timur. Nilai PDRB yang tinggi menunjukkan daerah dengan aktivitas perekonomian yang aktif, di mana penduduknya selalu bergerak. Sehingga kemungkinan terinfeksi COVID-19 dapat lebih tinggi apabila tidak dilakukan pengawasan.

Berdasarkan hasil prediksi dari model prediktif jumlah terkonfirmasi positif COVID-19 menggunakan *FBProphet* diperkirakan terus terjadi kenaikan jumlah positif COVID-19 di Jawa Timur hingga 30 Juli 2021. Model ini mampu memprediksi dengan baik karena hasil prediksi dapat mengikuti pergerakan dari data aktual dengan nilai RMSLE 0,25.

5.2 Saran

Analisis deskriptif pada penelitian ini terbatas hanya menggunakan dua belas atribut sosial, ekonomi dan demografi. Sedangkan dalam penyebaran kasus COVID-19 ada banyak faktor lainnya yang mempengaruhi seperti faktor cuaca, mobilitas masyarakat, serta kebijakan daerah setempat. Oleh karena itu, pada penelitian berikutnya faktor-faktor tersebut bisa diamati lebih lanjut. Pada analisis prediktif, variabel yang diperhatikan hanya penambahan jumlah positif COVID-19, belum mempertimbangkan adanya pengaruh dari variabel vaksinasi. Sehingga, pada penelitian berikutnya variabel vaksinasi bisa dimasukkan sebagai input pada model, untuk mendapatkan informasi yang lebih detail. Selain itu analisis prediktif dapat dilakukan menggunakan metode lain sebagai perbandingan.

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

DAFTAR PUSTAKA

- Aditya Satrio CB, Darmawan W, Nadia BU, Hanafiah N. 2021. Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. *Procedia Comput Sci.* 179(2020):524–532. doi:10.1016/j.procs.2021.01.036.
- Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, Rabczuk T, Atkinson PM. 2020. COVID-19 Outbreak Prediction with Machine Learning. *Algorithms.* 13(10):249. doi:10.3390/a13100249.
- Bank Dunia. 2012. Analisis Keuangan Publik Jawa Timur 2011: Mengoptimalkan Pengelolaan Keuangan Daerah untuk Pertumbuhan yang Inklusif.
- Basumatary P, Jain N. 2020. Linear Regression Model for Predictions of COVID-19 New Cases and New Deaths Based on May/June Data in Ethiopia. 11(10):1400–1406.
- [BPS] Badan Pusat Statistik. 2021. Provinsi Jawa Timur Dalam Angka [internet]. [diunduh 2021 Februari 26]. Tersedia pada: <https://jatim.bps.go.id/publication/2021/02/26/78c43a895e7f8ea378ffafc4/provinsi-jawa-timur-dalam-angka-2021.html>.
- Germaine. 2019. Time Series Cross-validation a Walk Forward Approach in Python. [Internet]. [diakses 2021 Jun 19]. Tersedia pada: <https://medium.com/eatpredlove/time-series-cross-validation-a-walkforward-approach-in-python-8534dd1db51a>.
- Gupta AK, Singh V, Mathur P, Travieso-Gonzalez CM. 2021. Prediction of COVID-19 pandemic measuring criteria using support vector machine, prophet and linear regression models in Indian scenario. *J Interdiscip Math.* 24(1):89–108. doi:10.1080/09720502.2020.1833458.
- Hale L. 2020. Why are COVID-19 numbers lower on the weekends?. [Internet]. [Diakses 2021 Jun 20]. Tersedia pada: <https://dearpandemic.org/covid-19-numbers-lower-on-the-weekends/>.
- Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, et al. 2020. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N Engl J Med.* 382(13):1199–1207. doi:10.1056/nejmoa2001316.
- Proietti T, Lutkepohl H. 2011. Munich Personal RePEc Archive Does the Box-Cox transformation help in forecasting macroeconomic time series ? Does the Box-Cox transformation help in forecasting macroeconomic time series ? *Munich Pers RePEc Arch.*(32294).
- Putri KR. 2017. Klasifikasi Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Indikator Kinerja Pembangunan Ekonomi Daerah dengan Metode Analisis Cluster. <http://repository.its.ac.id/41952/>.
- Qiu Y, Chen X, Shi W. 2020. Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. *J Popul Econ.* 33(4):1127–1172. doi:10.1007/s00148-020-00778-2.
- Rizal JG. Setelah Libur Panjang, Kasus Covid-19 Selalu Melonjak, Ini Datanya. [Internet]. [diakses 2021 Jun 9]. Tersedia pada: <https://www.kompas.com/trend/read/2020/12/29/120500865/setelah-libur-panjang-kasus-covid-19-selalu-melonjak-ini-datanya-?page=all>.

- Sannigrahi S, Pilla F, Basu B, Basu AS, Molter A. 2020. Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. *Sustain Cities Soc.* 62 July:102418. doi:10.1016/j.scs.2020.102418.
- Saraçlı S, Doğan N, Doğan I. 2013. Comparison of hierarchical cluster analysis methods by cophenetic correlation. *J Inequalities Appl.* 2013:1–8. doi:10.1186/1029-242X-2013-203.
- Satgas Covid-19] Satuan Tugas Penanganan COVID-19 Indonesia. 2021. [internet, diakses 2021 Mei 04]. Tersedia pada <https://covid19.go.id/peta-sebaran-covid19>.
- Saxena S. 2019. What's the Difference Between RMSE and RMSLE?. [Internet]. [diakses 2021 Jun 19]. Tersedia pada: <https://medium.com/analytics-vidhya/root-mean-square-log-error-rmse-vs-rmlse-935c6cc1802a>.
- Shereen MA, Khan S, Kazmi A, Bashir N, Siddique R. 2020. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *J Adv Res.* 24:91–98. doi:10.1016/j.jare.2020.03.005.
- Shinde GR, Kalamkar AB, Mahalle PN, Dey N, Chaki J, Hassanien AE. 2020. Forecasting Models for Coronavirus Disease (COVID-19): A Survey of the State-of-the-Art. *SN Comput Sci.* 1(4):1–15. doi:10.1007/s42979-020-00209-9.
- Shrestha N. 2020. Detecting Multicollinearity in Regression Analysis. *Am J Appl Math Stat.* 8(2):39–42. doi:10.12691/ajams-8-2-1.
- Taylor SJ, Letham B. 2018. Forecasting at Scale. *Am Stat.* 72(1):37–45. doi:10.1080/00031305.2017.1380080.
- Tosepu R, Gunawan J, Effendy DS, Ahmad LOAI, Lestari H, Bahar H, Asfian P. 2020. Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *Sci Total Environ.* 725. doi:10.1016/j.scitotenv.2020.138436.
- Ulinuha N, Veriani R. 2020. Analisis Cluster dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage *J Nas Inform dan* 1. <https://jurnal.uisu.ac.id/index.php/infotekjar/article/view/2464>.

LAMPIRAN

Lampiran 1 Tabel data sebelum standarisasi

Kab/Kota	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Kab Bangkalan	3	110	64,11	20,56	17.514,62	791.935	8,77	1059	1.060.377	30,84	97,24	12,83
Kab Banyuwangi	12	316	70,62	8,06	53.295,11	990.798	5,34	295	1.708.114	60,21	100,27	15,84
Kab Blitar	8	184	70,58	9,33	24.945,46	885.475	3,82	916	1.223.745	45	101,53	17,11
Kab Bojonegoro	9	221	69,04	12,87	69.703,42	904.017	4,92	592	1.301.635	25,92	100,89	14,75
Kab Bondowoso	3	88	66,43	14,17	13.451,77	807.746	4,13	509	776.151	39,51	97,03	16,35
Kab Gresik	18	297	76,11	12,4	97.616,6	1304.981	8,21	1101	1.311.215	66,95	101,44	9,79
Kab Jember	10	282	67,11	10,09	52.586,56	875.962	5,12	820	2.536.729	50,8	99,47	14,3
Kab Jombang	11	248	72,97	9,94	27.657,58	1.000.946	7,48	1182	1.318.062	69,22	101,71	13,26
Kab Kediri	9	256	72,05	11,4	28.490,95	897.042	5,24	1180	1.635.294	52,54	102,03	14,5
Kab Lamongan	10	219	72,58	13,85	26.972,65	1.018.645	5,13	754	1.344.165	27,9	100,18	14,66
Kab Lumajang	7	141	65,46	9,83	21.933,79	756.384	3,36	625	1.119.251	35,11	97,86	14,62
Kab Madiun	3	104	71,73	11,46	12.939,58	917.157	4,8	717	744.350	36,36	98,26	17,9
Kab Magetan	2	95	73,92	10,35	13.020,89	941.723	3,74	974	670.812	45,85	96,84	20,31
Kab Malang	20	345	70,36	10,15	66.545,47	957.860	5,49	752	2.654.448	54,82	101,61	14,2
Kab Mojokerto	11	192	73,83	10,57	57.818,42	1.206.234	5,75	1559	1.119.209	55,46	101,2	11,83
Kab Nganjuk	5	202	71,72	11,62	17.990,36	880.817	4,8	902	1.103.902	43,14	101,21	15,51
Kab Ngawi	3	181	70,54	15,44	13.479,74	817.267	5,44	671	870.057	25,3	98,6	17,48
Kab Pacitan	2	89	68,39	14,54	10.837,87	789.088	2,28	422	586.110	21,77	100,45	20,25
Kab Pamekasan	6	133	66,26	14,6	11.117,62	714.598	3,49	1073	850.057	30,46	96,27	10,78
Kab Pasuruan	8	184	68,6	9,26	103.152,8	1.014.042	6,24	1090	1.605.969	50,9	100,19	10,14
Kab Ponorogo	6	147	70,81	9,95	14.168,62	1.018.283	4,45	727	949.318	34,73	99,83	19,23
Kab Probolinggo	5	110	66,07	18,61	22.898,24	859.198	4,86	679	1.152.537	40,91	97,4	13,04

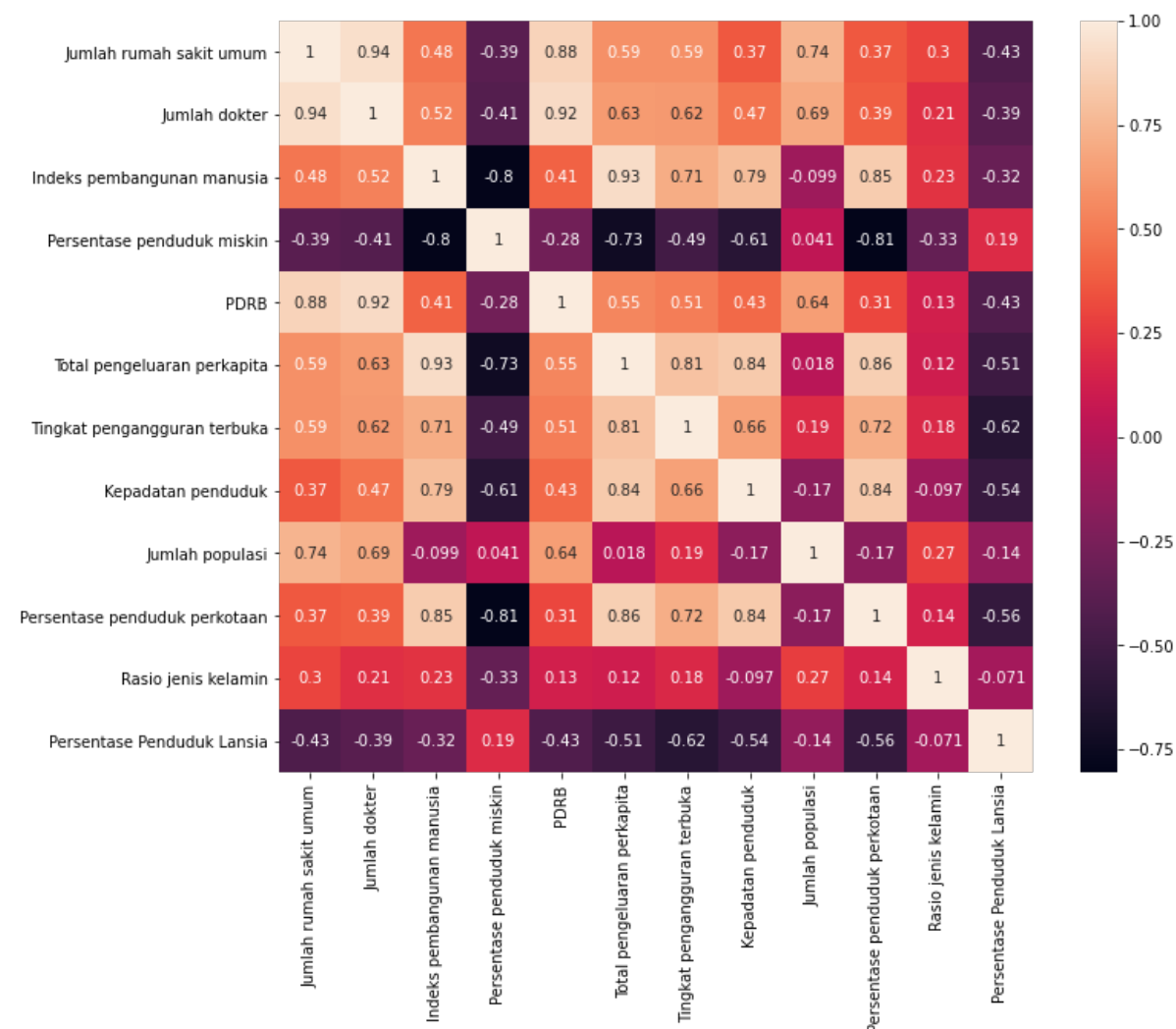
Kab Sampang	3	79	62,7	22,78	13.953,74	667.972	3,35	786	969.694	20,06	98,7	10,51
Kab Sidoarjo	21	589	80,29	5,59	135.305,3	1.602.881	10,97	3283	2.082.801	98,43	101,39	8,19
Kab Situbondo	5	96	67,38	12,22	13.282,84	773.896	3,85	411	685.967	51,47	96,43	14,67
Kab Sumenep	3	88	66,43	20,18	23.546,51	893.067	2,84	563	1.124.436	23,41	93,3	14,44
Kab Trenggalek	2	98	69,74	11,62	12.502,39	863.724	4,11	637	731.125	35,78	100,99	17,23
Kab Tuban	5	143	68,4	15,91	42.705,01	890.624	4,81	653	1.198.012	30,99	99,78	13,25
Kab Tulungagung	9	231	73	7,33	26.455,76	1.032.860	4,61	1032	1.089.775	52,95	100,13	15,94
Kota Batu	5	104	75,9	3,89	11.025,81	1.369.206	5,93	1558	213.046	97,04	101,47	12,43
Kota Blitar	5	125	78,57	7,78	4.722,55	1.369.646	6,68	4579	149.149	100	99,05	13,44
Kota Kediri	8	195	78,23	7,69	84.374,98	1.249.789	6,21	4524	286.796	100	100,3	11,44
Kota Madiun	6	155	80,91	4,98	10.262,44	1.569.951	8,32	5754	195.175	100	95,7	14,74
Kota Malang	13	438	81,45	4,44	51.154,53	1.743.568	9,61	5808	843.810	100	99,05	11,04
Kota Mojokerto	6	111	78,04	6,24	4.801,46	1.383.231	6,74	6553	132.434	100	98,41	10,96
Kota Pasuruan	2	96	75,26	6,66	5.706,6	1.209.888	6,33	5894	208.006	100	100,08	9,46
Kota Probolinggo	2	131	73,27	7,43	8.035,27	1.213.358	6,7	4229	239.649	97,89	98,35	10,66
Kota Surabaya	37	1084	82,23	5,02	390.936,4	1.879.145	9,79	8200	2.874.314	100	98,35	9,16

dengan:

X1 = Jumlah rumah sakit umum
X2 = Jumlah dokter
X3 = Indeks pembangunan manusia
X4 = Persentase penduduk miskin
X5 = PDRB
X6 = Total pengeluaran perkapita

X7 = Tingkat pengangguran terbuka
X8 = Kepadatan penduduk
X9 = Jumlah populasi
X10 = Persentase penduduk perkotaan
X11 = Rasio jenis kelamin
X12 = Persentase Penduduk Lansia

Lampiran 2 *Heatmap* koefisien korelasi antar atribut



Lampiran 3 Nilai *p-value* korelasi antar atribut

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
X1	-											
X2	0,00	-										
X3	0,00	0,00	-									
X4	0,02	0,01	0,00	-								
X5	0,00	0,00	0,01	0,09	-							
X6	0,00	0,00	0,00	0,00	0,00	-						
X7	0,00	0,00	0,00	0,00	0,00	0,00	-					
X8	0,02	0,00	0,00	0,00	0,01	0,00	0,00	-				
X9	0,00	0,00	0,55*	0,81*	0,00	0,91*	0,25	0,31	-			
X10	0,02	0,02	0,00	0,00	0,06	0,00	0,00	0,00	0,30	-		
X11	0,07	0,20	0,16	0,04	0,43	0,49	0,28	0,56*	0,11	0,40	-	
X12	0,01	0,02	0,05	0,25	0,01	0,00	0,00	0,00	0,40	0,00	0,67*	-

*tidak berkorelasi nyata pada taraf 5%

Lampiran 4 Nilai *Variance Inflation Factor*

Variabel	VIF pada iterasi ke-i					
	1	2	3	4	5	6
Jumlah dokter	21,492*	-	-	-	-	-
Total pengeluaran perkapita	18,950	18,947*	-	-	-	-
Indeks pembangunan manusia	16,987	16,829	10,509	7,872	7,437*	-
Jumlah rumah sakit umum	16,815	14,188	14,017*	-	-	-
Persentase penduduk perkotaan	13,276	12,910	12,903	12,902*	-	-
PDRB	8,851	6,301	6,109	4,035	3,813	3,513
Jumlah populasi	8,385	7,294	7,219	3,624	3,569	3,160
Kepadatan penduduk	7,457	6,531	6,485	6,308	5,566	4,701
Persentase penduduk miskin	6,032	5,959	5,799	5,791	3,262	2,414
Tingkat pengangguran terbuka	4,825	4,433	3,860	3,847	3,636	2,609
Persentase penduduk lansia	3,916	3,722	3,549	3,274	2,332	1,972
Rasio jenis kelamin	1,928	1,925	1,701	1,699	1,685	1,547

* variabel yang dihilangkan pada setiap iterasi, karena nilai VIF paling besar

Lampiran 5 Jarak *Euclidean* antar kabupaten/kota di Jawa Timur berdasarkan data sosial, ekonomi dan demografi

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
0	0	3,93	4,39	3,37	3,02	3,25	3,95	3,36	3,78	2,9	3,7	3,38	4,25	4,43	3,44	3,58	2,66	4,67	3,11
1	3,93	0	1,44	1,37	2,7	2,89	1,51	1,81	1,37	1,53	1,96	2,17	3,01	1,77	1,87	1,47	2,39	3,15	3,48
2	4,39	1,44	0	1,47	2,64	3,53	2,56	2,26	1,39	1,61	2,06	1,96	2,74	2,6	2,11	0,92	2,25	2,09	3,63
3	3,37	1,37	1,47	0	2,39	2,4	2,12	1,74	1,1	0,79	2,01	2,1	3,15	2,2	1,31	0,96	1,93	2,69	3,03
4	3,02	2,7	2,64	2,39	0	4,06	3,25	3,35	3,05	1,98	1,37	1,07	1,59	3,99	2,99	2,29	1,15	2,37	1,93
5	3,25	2,89	3,53	2,4	4,06	0	3,14	1,71	2,49	2,59	3,69	3,9	5,04	2,94	1,64	2,86	3,61	4,94	3,89
6	3,95	1,51	2,56	2,12	3,25	3,14	0	2,53	1,95	2,06	2,51	3,11	3,81	1,13	2,51	2,46	3,08	4,04	3,55
7	3,36	1,81	2,26	1,74	3,35	1,71	2,53	0	1,35	1,74	2,92	2,86	4,01	2,37	1,19	1,65	2,74	3,89	3,73
8	3,78	1,37	1,39	1,1	3,05	2,49	1,95	1,35	0	1,19	2,48	2,62	3,66	1,71	1,38	1,01	2,5	3,11	3,57
9	2,9	1,53	1,61	0,79	1,98	2,59	2,06	1,74	1,19	0	1,76	1,81	2,92	2,36	1,5	0,88	1,49	2,63	2,63
10	3,7	1,96	2,06	2,01	1,37	3,69	2,51	2,92	2,48	1,76	0	1,49	2,08	3,27	2,39	1,91	1,96	2,69	1,9
11	3,38	2,17	1,96	2,1	1,07	3,9	3,11	2,86	2,62	1,81	1,49	0	1,23	3,69	2,73	1,78	0,98	2	2,74
12	4,25	3,01	2,74	3,15	1,59	5,04	3,81	4,01	3,66	2,92	2,08	1,23	0	4,51	3,84	2,86	1,95	2,2	3,3
13	4,43	1,77	2,6	2,2	3,99	2,94	1,13	2,37	1,71	2,36	3,27	3,69	4,51	0	2,5	2,56	3,58	4,3	4,33
14	3,44	1,87	2,11	1,31	2,99	1,64	2,51	1,19	1,38	1,5	2,39	2,73	3,84	2,5	0	1,49	2,68	3,63	3,03
15	3,58	1,47	0,92	0,96	2,29	2,86	2,46	1,65	1,01	0,88	1,91	1,78	2,86	2,56	1,49	0	1,77	2,3	3,12
16	2,66	2,39	2,25	1,93	1,15	3,61	3,08	2,74	2,5	1,49	1,96	0,98	1,95	3,58	2,68	1,77	0	2,12	2,7
17	4,67	3,15	2,09	2,69	2,37	4,94	4,04	3,89	3,11	2,63	2,69	2	2,2	4,3	3,63	2,3	2,12	0	3,85
18	3,11	3,48	3,63	3,03	1,93	3,89	3,55	3,73	3,57	2,63	1,9	2,74	3,3	4,33	3,03	3,12	2,7	3,85	0
19	3,65	2,14	3	2,01	3,55	1,45	2,23	1,9	2,18	2,26	2,8	3,45	4,44	2,31	1,34	2,55	3,42	4,58	3,26
20	4,07	1,84	1,25	1,99	1,97	4,03	2,99	2,74	2,26	1,89	1,92	1,03	1,66	3,34	2,75	1,49	1,57	1,67	3,51
21	2,04	3,07	3,27	2,37	1,64	3,39	3,08	3,21	2,98	1,87	2,17	2,38	3,17	3,73	2,78	2,61	1,81	3,36	1,56
22	2,99	4,13	3,99	3,1	2,88	3,91	4,05	4	3,6	2,74	3,22	3,6	4,38	4,45	3,35	3,32	3,01	3,88	2,21
23	5,08	4,37	5,32	4,56	6,07	2,7	4,27	3,46	4,32	4,75	5,49	5,78	6,71	4	3,73	4,91	5,69	7,07	5,83
24	3,25	2,85	2,91	2,65	0,79	4,07	3,36	3,47	3,29	2,28	1,15	1,51	1,93	4,17	2,99	2,57	1,82	2,9	1,46

25	3,65	4,74	4,92	4,36	2,54	5,58	4,57	5,39	5,04	3,94	3,27	3,56	3,53	5,5	4,89	4,55	3,35	4,37	2,37
26	3,93	1,99	0,98	1,56	2,1	3,6	3,13	2,41	1,85	1,52	1,98	1,44	2,36	3,3	2,22	0,89	1,64	1,55	3,29
27	2,64	2,13	2,21	1,1	1,96	2,52	2,45	2,16	1,75	0,78*	1,9	2,14	3,21	2,78	1,62	1,47	1,69	2,88	2,21
28	4,03	1,16	1,03	1,56	2,27	3,23	2,41	2	1,66	1,58	1,5	1,58	2,43	2,74	1,81	1,12	2,11	2,59	3,13
29	4,7	2,92	2,73	2,94	3,7	3,2	4,05	2,33	2,86	3,03	3,06	3,14	4,01	4,11	2,16	2,5	3,56	4,06	3,92
30	3,85	3,44	3,33	3,29	3,22	3,54	4,23	2,85	3,38	3,11	3,06	2,86	3,54	4,59	2,65	2,92	3,14	4,14	3,41
31	4,2	3,36	3,37	3	3,74	2,81	4,09	2,71	3,19	3,15	3,31	3,44	4,24	4,23	2,12	2,98	3,63	4,48	3,63
32	4,45	4,56	4,83	4,77	4,06	4,84	5,12	4,32	4,94	4,53	4,09	3,82	4,04	5,72	4,29	4,53	4,18	5,46	4,28
33	4,44	4,12	4,62	4,28	4,83	3,37	4,5	3,25	4,17	4,23	4,45	4,47	5,19	4,71	3,36	4,22	4,63	5,98	4,71
34	4,57	4,41	4,41	4,3	4,23	4,13	4,93	3,77	4,31	4,13	3,94	4,04	4,6	5,31	3,45	4,03	4,3	5,3	3,93
35	4,63	4,21	4,13	3,98	4,36	3,62	4,75	3,35	3,88	3,88	3,91	4,14	4,89	4,96	2,94	3,71	4,34	5,21	3,92
36	3,78	3,66	3,77	3,5	3,43	3,34	4,25	3	3,62	3,31	3,12	3,29	4,03	4,67	2,64	3,28	3,53	4,71	3,07
37	8,15	7,48	8,35	7,6	8,72	6,59	7,07	7,45	7,74	7,95	8,23	8,6	9,05	7,03	7,22	8,22	8,59	9,68	8,45

(lanjutan matriks jarak)

	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
0	3,65	4,07	2,04	2,99	5,08	3,25	3,65	3,93	2,64	4,03	4,7	3,85	4,2	4,45	4,44	4,57	4,63	3,78	8,15
1	2,14	1,84	3,07	4,13	4,37	2,85	4,74	1,99	2,13	1,16	2,92	3,44	3,36	4,56	4,12	4,41	4,21	3,66	7,48
2	3	1,25	3,27	3,99	5,32	2,91	4,92	0,98	2,21	1,03	2,73	3,33	3,37	4,83	4,62	4,41	4,13	3,77	8,35
3	2,01	1,99	2,37	3,1	4,56	2,65	4,36	1,56	1,1	1,56	2,94	3,29	3	4,77	4,28	4,3	3,98	3,5	7,6
4	3,55	1,97	1,64	2,88	6,07	0,79	2,54	2,1	1,96	2,27	3,7	3,22	3,74	4,06	4,83	4,23	4,36	3,43	8,72
5	1,45	4,03	3,39	3,91	2,7	4,07	5,58	3,6	2,52	3,23	3,2	3,54	2,81	4,84	3,37	4,13	3,62	3,34	6,59
6	2,23	2,99	3,08	4,05	4,27	3,36	4,57	3,13	2,45	2,41	4,05	4,23	4,09	5,12	4,5	4,93	4,75	4,25	7,07
7	1,9	2,74	3,21	4	3,46	3,47	5,39	2,41	2,16	2	2,33	2,85	2,71	4,32	3,25	3,77	3,35	3	7,45
8	2,18	2,26	2,98	3,6	4,32	3,29	5,04	1,85	1,75	1,66	2,86	3,38	3,19	4,94	4,17	4,31	3,88	3,62	7,74
9	2,26	1,89	1,87	2,74	4,75	2,28	3,94	1,52	0,78	1,58	3,03	3,11	3,15	4,53	4,23	4,13	3,88	3,31	7,95
10	2,8	1,92	2,17	3,22	5,49	1,15	3,27	1,98	1,9	1,5	3,06	3,06	3,31	4,09	4,45	3,94	3,91	3,12	8,23

11	3,45	1,03	2,38	3,6	5,78	1,51	3,56	1,44	2,14	1,58	3,14	2,86	3,44	3,82	4,47	4,04	4,14	3,29	8,6
12	4,44	1,66	3,17	4,38	6,71	1,93	3,53	2,36	3,21	2,43	4,01	3,54	4,24	4,04	5,19	4,6	4,89	4,03	9,05
13	2,31	3,34	3,73	4,45	4	4,17	5,5	3,3	2,78	2,74	4,11	4,59	4,23	5,72	4,71	5,31	4,96	4,67	7,03
14	1,34	2,75	2,78	3,35	3,73	2,99	4,89	2,22	1,62	1,81	2,16	2,65	2,12	4,29	3,36	3,45	2,94	2,64	7,22
15	2,55	1,49	2,61	3,32	4,91	2,57	4,55	0,89	1,47	1,12	2,5	2,92	2,98	4,53	4,22	4,03	3,71	3,28	8,22
16	3,42	1,57	1,81	3,01	5,69	1,82	3,35	1,64	1,69	2,11	3,56	3,14	3,63	4,18	4,63	4,3	4,34	3,53	8,59
17	4,58	1,67	3,36	3,88	7,07	2,9	4,37	1,55	2,88	2,59	4,06	4,14	4,48	5,46	5,98	5,3	5,21	4,71	9,68
18	3,26	3,51	1,56	2,21	5,83	1,46	2,37	3,29	2,21	3,13	3,92	3,41	3,63	4,28	4,71	3,93	3,92	3,07	8,45
19	0	3,57	3,12	3,82	3,01	3,42	5,01	3,29	2,26	2,55	2,99	3,41	2,67	4,59	3,4	3,95	3,53	3,13	6,33
20	3,57	0	3,08	4,12	5,8	2,4	4,39	1,03	2,45	1,28	3,08	3,2	3,63	4,33	4,7	4,4	4,39	3,76	8,67
21	3,12	3,08	0	1,63	5,57	1,83	2,38	2,86	1,38	3,02	4,18	3,64	3,85	4,62	4,82	4,43	4,38	3,57	8,32
22	3,82	4,12	1,63	0	6,29	2,99	3,1	3,56	2,07	4	4,8	4,45	4,42	5,77	5,73	5,08	4,8	4,3	9,01
23	3,01	5,8	5,57	6,29	0	5,99	7,43	5,67	4,88	4,87	4,58	4,73	3,98	5,3	3,08	4,79	4,45	4,44	5,04
24	3,42	2,4	1,83	2,99	5,99	0	2,52	2,47	2,19	2,35	3,52	3,1	3,59	3,87	4,67	3,98	4,12	3,12	8,64
25	5,01	4,39	2,38	3,1	7,43	2,52	0	4,51	3,59	4,6	5,93	5,09	5,5	5,37	6,35	5,66	5,93	4,98	9,26
26	3,29	1,03	2,86	3,56	5,67	2,47	4,51	0	1,96	1,31	2,68	3,05	3,29	4,61	4,69	4,25	4,03	3,55	8,77
27	2,26	2,45	1,38	2,07	4,88	2,19	3,59	1,96	0	2,14	3,3	3,27	3,17	4,7	4,41	4,2	3,91	3,32	7,93
28	2,55	1,28	3,02	4	4,87	2,35	4,6	1,31	2,14	0	2,17	2,67	2,82	4	3,91	3,73	3,56	3,01	7,98
29	2,99	3,08	4,18	4,8	4,58	3,52	5,93	2,68	3,3	2,17	0	2,15	2,12	3,83	3,26	2,98	2,49	2,29	8,26
30	3,41	3,2	3,64	4,45	4,73	3,1	5,09	3,05	3,27	2,67	2,15	0	1,56	2,11	2,31	1,34	1,58	1,01	7,77
31	2,67	3,63	3,85	4,42	3,98	3,59	5,5	3,29	3,17	2,82	2,12	1,56	0	3,12	2,29	1,9	1,55	1,58	6,78
32	4,59	4,33	4,62	5,77	5,3	3,87	5,37	4,61	4,7	4	3,83	2,11	3,12	0	2,48	2,08	3,01	2,26	7,61
33	3,4	4,7	4,82	5,73	3,08	4,67	6,35	4,69	4,41	3,91	3,26	2,31	2,29	2,48	0	2,04	2,23	2,15	6,22
34	3,95	4,4	4,43	5,08	4,79	3,98	5,66	4,25	4,2	3,73	2,98	1,34	1,9	2,08	2,04	0	1,06	1,17	7,49
35	3,53	4,39	4,38	4,8	4,45	4,12	5,93	4,03	3,91	3,56	2,49	1,58	1,55	3,01	2,23	1,06	0	1,28	7,53
36	3,13	3,76	3,57	4,3	4,44	3,12	4,98	3,55	3,32	3,01	2,29	1,01	1,58	2,26	2,15	1,17	1,28	0	7,57
37	6,33	8,67	8,32	9,01	5,04	8,64	9,26	8,77	7,93	7,98	8,26	7,77	6,78	7,61	6,22	7,49	7,53	7,57	0

* jarak paling kecil

Keterangan:

0	= Kab. Bangkalan	19	= Kab. Pasuruan
1	= Kab. Banyuwangi	20	= Kab. Ponorogo
2	= Kab. Blitar	21	= Kab. Probolinggo
3	= Kab. Bojonegoro	22	= Kab. Sampang
4	= Kab. Bondowoso	23	= Kab. Sidoarjo
5	= Kab. Gresik	24	= Kab. Situbondo
6	= Kab. Jember	25	= Kab. Sumenep
7	= Kab. Jombang	26	= Kab. Trenggalek
8	= Kab. Kediri	27	= Kab. Tuban
9	= Kab. Lamongan	28	= Kab. Tulungagung
10	= Kab. Lumajang	29	= Kota Batu
11	= Kab. Madiun	30	= Kota Blitar
12	= Kab. Magetan	31	= Kota Kediri
13	= Kab. Malang	32	= Kota Madiun
14	= Kab. Mojokerto	33	= Kota Malang
15	= Kab. Nganjuk	34	= Kota Mojokerto
16	= Kab. Ngawi	35	= Kota Pasuruan
17	= Kab. Pacitan	36	= Kota Probolinggo
18	= Kab. Pamekasan	37	= Kota Surabaya

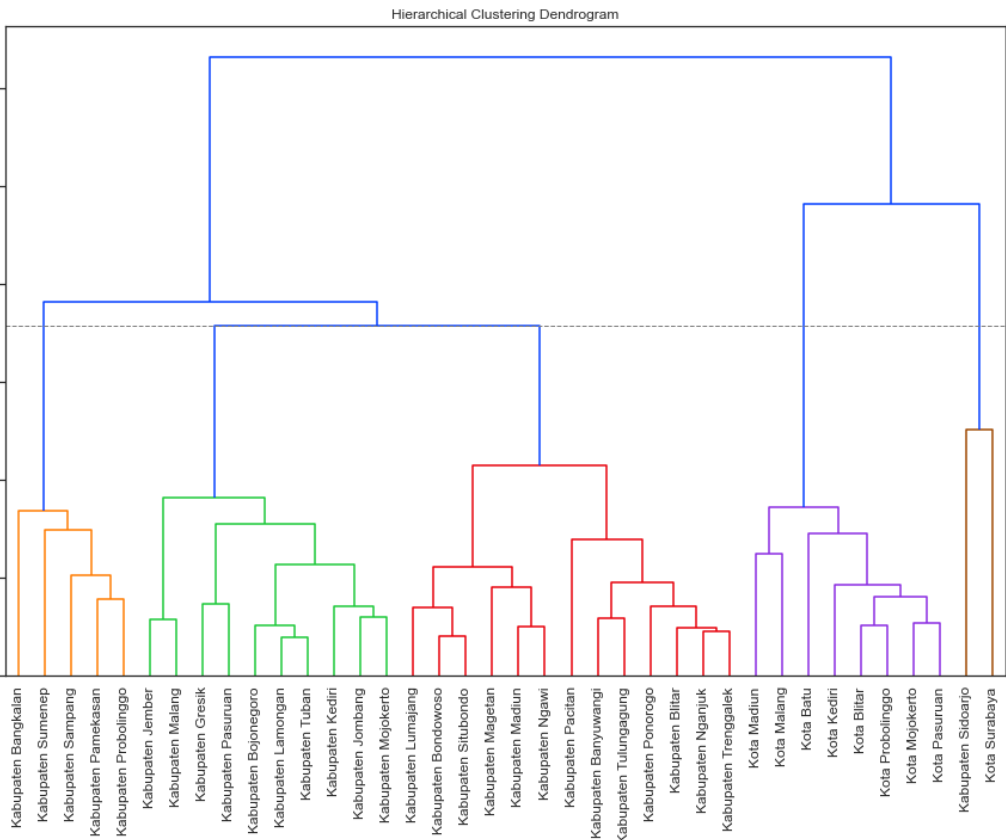
- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Lampiran 6 Matriks pautan pada metode hirarki dengan pautan *Ward*

Penggabungan ke-	Cluster P	Cluster Q	jarak	jumlah_anggota_cluster
0	9	27	0.775761	2
1	4	24	0.79142	2
2	15	26	0.893122	2
3	2	40	0.973385	3
4	11	16	0.984907	2
5	30	36	1.009737	2
6	3	38	1.015299	3
7	34	35	1.059368	2
8	6	13	1.133915	2
9	1	28	1.158986	2
10	7	14	1.188765	2
11	10	39	1.386097	3
12	20	41	1.407391	4
13	8	48	1.415474	3
14	5	19	1.448356	2
15	18	21	1.559214	2
16	43	45	1.601392	4
17	12	42	1.793312	3
18	31	54	1.857915	5
19	47	50	1.901779	6
20	22	53	2.052302	3
21	49	55	2.207344	6
22	44	51	2.266883	6
23	32	33	2.481333	2
24	17	57	2.783406	7
25	29	56	2.909222	6
26	25	58	2.965695	4
27	52	60	3.105335	8
28	0	64	3.370172	5
29	61	63	3.438603	8
30	46	65	3.631315	10
31	59	62	4.289852	13
32	23	37	5.036887	2
33	68	69	7.157937	23
34	66	71	7.644557	28
35	67	70	9.639474	10
36	72	73	12.64091	38

Lampiran 7 Dendrogram hasil pengelompokan menggunakan metode *Agglomerative Hierarchical Clustering*

@Hak cipta milik IPB University



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

RIWAYAT HIDUP

Penulis dilahirkan di Kabupaten Limapuluh Kota, Sumatera Barat pada 4 Juni 1999. Pada tahun 2017, penulis lulus dari SMAN 1 Suliki dan diterima melalui jalur SNMPTN sebagai mahasiswa program sarjana (S-1) di Departemen Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, IPB.

Selama mengikuti program S-1, penulis aktif pada beberapa organisasi kampus. Pada tahun 2019 penulis tergabung pada BEM FMIPA IPB sebagai Sekretaris Departemen Sosial dan Masyarakat. Kemudian pada tahun 2020 penulis bergabung dengan BEM KM IPB sebagai Staff Biro Riset dan Pengembangan.

@Hak cipta milik IPB University

- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.