

ANALISIS KETERCAPAIAN VAKSINASI TERHADAP PENYEBARAN COVID-19

Ferdy Berliano Putra
Syafrial Fachri Pane

Jl. Sari Asih No.54, Sarijadi, Kec. Sukasari, Kota Bandung,
Jawa Barat 40151

D4 Teknik Informatika, Politeknik Pos Indonesia

Jalan Sari Asih No.54, Kota Bandung, Jawa Barat

Judul : Analisis Ketercapaian Vaksinasi Terhadap Penyebaran COVID-19

(Analisis Ketercapaian Vaksinasi Terhadap Penyebaran COVID-19)
2022,

... Halaman, 16 cm x 23 cm

Penulis

Ferdy Berliano Putra

Syafrial Fachri Pane, S.T.,M.T.I.,EBDP

Penyunting : Nama Penguji

Layout : Ferdy Berliano Putra

Sanksi Pelanggaran Pasal 113

Undang-Undang Nomor 28 Tahun 2014

tentang Hak Cipta

- (1) Setiap Orang yang dengan tanpa hak melakukan pelanggaran hak ekonomi sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf i untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 1 (satu) tahun dan/atau pidana denda paling banyak Rp100.000.000 (seratus juta rupiah).
- (2) Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf e, dan/atau huruf h untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).
- (3) Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf a, huruf b, huruf c, dan/atau huruf d untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 4 (empat) tahun dan/atau pidana denda paling banyak Rp1.000.000.000,00 (satu miliar rupiah).
- (4) Setiap Orang yang memenuhi unsur sebagaimana dimaksud pada ayat (3) yang dilakukan dalam bentuk pembajakan, dipidana dengan pidana penjara paling lama 10 (sepuluh) tahun dan/atau pidana denda paling banyak Rp4.000.000.000,00 (empat miliar rupiah).

Kata Sambutan

Puji syukur penulis panjatkan kehadiran Allah SWT atas terbitnya buku dengan judul “Analisis Ketercapaian Vaksinasi Terhadap Penyebaran COVID-19”. Buku ini disusun dalam konteks mendorong masyarakat khususnya mahasiswa yang sedang mencari referensi kebutuhan buku laporan, baik itu buku laporan magang ataupun buku laporan tugas akhir. Semoga dengan adanya buku ini dapat memberikan banyak manfaat kepada kalian.

Kata Pengantar

Puji syukur penulis panjatkan kehadiran Allah SWT karen berkat rahmat dan hidayah-Nya lah dapat menyelesaikan Internship 1 berupa pembuatan sistem yang berjudul “Analisis Ketercapaian Vaksinasi Terhadap Penyebaran COVID-19” ini dapat selesai tepat waktunya sebagai syarat kelulusan matakuliah Internship 1.

Dalam kesempatan ini penulis tidak lupa menyampaikan ucapan terima kasih kepada semuapihak yang telah memberikan bantuan moral dan spiritual, langsung maupun tidak langsung dalam menyelesaikan laporan ini, terutama kepada:

1. Kedua orang tua tercinta yang senantiasa memberikan dukungan dan doanya.
2. Dr. Ir. Agus Purnomo, M.T. selaku Direktur Politeknik Pos Indonesia.
3. M.Yusril Helmi Setyawan, S. Kom., M. Kom. selaku Ketua Program Studi DIV Teknik Informatika.
4. Cahyo Prianto, S.Pd., M.T selaku koordinator Internship 1.
5. Syafrial Fachri Pane, S.T., M.T.I., EBDP.CDSP,SFPC selaku Pembimbing Internal Internship 1 yang telah memberikan pengarahan dan membimbing kepada penulis dan juga selaku dosen wali kelas D4 TI 4A.
6. Serta semua pihak yang tidak dapat penulis sebutkan satu persatu.

Daftar Isi

KATA SAMBUTAN.....	III
KATA PENGANTAR	IV
DAFTAR ISI.....	V
DAFTAR GAMBAR.....	VII
DAFTAR TABEL	IX
BAB I PENDAHULUAN	I-1
1.1 LATAR BELAKANG.....	I-1
1.2 IDENTIFIKASI MASALAH.....	I-3
1.3 TUJUAN	I-3
1.4 MANFAAT	I-3
1.5 RUANG LINGKUP	I-4
BAB II LANDASAN TEORI.....	II-5
2.1 PENELITIAN TERDAHULU	II-5
2.1.1 <i>Penelitian Terkait</i>	II-5
2.2 LANDASAN TEORI	II-11
2.2.1 <i>Vaksinasi</i>	II-11
2.2.2 <i>Regresi</i>	II-12
2.2.3 <i>AdaBoost or Adaptive Boosting Regressor</i>	II-13
2.2.4 <i>Regressor Model Evaluation dan Mean Absolute Error</i>	II-13
2.2.5 <i>Python</i>	II-14
2.2.6 <i>Matplotlib</i>	II-14
2.2.7 <i>Seaborn</i>	II-14
2.2.8 <i>Pandas</i>	II-14
2.2.9 <i>Numpy</i>	II-15
2.2.10 <i>Scikit-learn</i>	II-15
2.2.11 <i>Outlier</i>	II-15
2.2.12 <i>Interpolasi Linier</i>	II-16
2.3 METODOLOGI PENELITIAN	II-16
BAB III METODOLOGI PENELITIAN	III-24
3.1 METODE PENELITIAN.....	III-24
3.2 TAHAPAN-TAHAPAN DIAGRAM ALUR	III-25

3.2.1	<i>Business Understanding</i>	III-25
3.2.2	<i>Data Understanding</i>	III-26
3.2.3	<i>Data preparation</i>	III-26
3.2.4	<i>Modelling</i>	III-27
3.2.5	<i>Evaluation</i>	III-27
BAB IV	ANALISIS DAN HASIL PEMBAHASAN	IV-28
4.1	ANALISIS	IV-28
4.1.1	<i>Data yang digunakan</i>	IV-28
4.2	HASIL PEMBAHASAN	IV-30
BAB V	KESIMPULAN DAN SARAN	V-48
5.1	KESIMPULAN	V-48
5.2	SARAN	V-48
DAFTAR PUSTAKA		1

Daftar Gambar

Gambar 2. 1 Grafik intepolasi linier	II-16
Gambar 2. 2 Alur diagram CRISP-DM	II-17
Gambar 2. 3 Langkah Diagram CRISP-DM.....	II-19
Gambar 3. 1 Alur Diagram Metodologi Penelitian	III-25
Gambar 4. 1 Tampilan data.....	IV-29
Gambar 4. 2 Tampilan Data Head.....	IV-29
Gambar 4. 3 Tampilan Data Tail.....	IV-30
Gambar 4. 4 Langkah Business Understanding.....	IV-30
Gambar 4. 5 Langkah Data Understanding.....	IV-31
Gambar 4. 6 Kumpulan data perbulan	IV-32
Gambar 4. 7 Kumpulan Data perhari.....	IV-32
Gambar 4. 8 Data dalam bentuk Excel	IV-33
Gambar 4. 9 Melihat nilai kosong pada data.....	IV-34
Gambar 4. 10 Data Outlier Tabel Dosis 1	IV-34
Gambar 4. 11 Data Outlier Tabel Dosis 2	IV-35
Gambar 4. 12 Data Outlier Tabel Total Vaksin Diberikan.....	IV-35
Gambar 4. 13 Data Outlier Tabel Sasaran.....	IV-36
Gambar 4. 14 Data Outlier Tabel Belum Vaksin	IV-36
Gambar 4. 15 Langkah Data Preparation	IV-37
Gambar 4. 16 Seleksi Data.....	IV-38
Gambar 4. 17 Interpolasi	IV-39
Gambar 4. 18 Pembersihan Outlier Tabel Dosis1 Menggunakan Metode IQR	IV-39
Gambar 4. 19 Pembersihan Outlier Tabel Dosis2 Menggunakan Metode IQR	IV-39
Gambar 4. 20 Pembersihan Outlier Tabel Total Vaksin diberikan Menggunakan Metode IQR	IV-40
Gambar 4. 21 Data Oulier Setelah dibersihkan Pada Tabel Dosis1	IV-40
Gambar 4. 22 Data Oulier Setelah dibersihkan Pada Tabel Dosis2	IV-41

Gambar 4. 23 Data Oulier Setelah dibersihkan Pada Tabel Total Vaksin diberikan.....	IV-41
Gambar 4. 24 Data Oulier Setelah dibersihkan Pada Tabel Sasaran	IV-42
Gambar 4. 25 Data Oulier Setelah dibersihkan Pada Tabel Belum Vaksin	IV-42
Gambar 4. 26 Mengubah Data String Menjadi Data Numerik	IV-43
Gambar 4. 27 Langkah Modelling	IV-43
Gambar 4. 28 Split Data	IV-44
Gambar 4. 29 Penggunaan Model	IV-44
Gambar 4. 30 Nilai Model dan Korelasi.....	IV-44
Gambar 4. 31 Visualisais Algoritma AdaBoost Regressor	IV-45
Gambar 4. 32 Perkembangan vaksinasi Total Dosis yang Sudah Diberikan	IV-45
Gambar 4. 33 Perkembangan Vaksinasi Dosis1 yang Sudah Diberikan .	IV-46
Gambar 4. 34 Perkembangan Vaksinasi Dosis2 yang Sudah Diberikan .	IV-46
Gambar 4. 35 Nilai MAE dan RMSE.....	IV-47

Daftar Tabel

Tabel 2. 1 Penelitian terkait.....	II-5
------------------------------------	------

Bab I

Pendahuluan

1.1 Latar Belakang

Virus bukanlah suatu hal yang baru. Salah satu contoh virus yang mempengaruhi banyak orang adalah cacar air namun lain halnya dengan COVID-19. Perbedaan antara bakteri dan virus terletak pada tingkatan biologis namun sama membuat manusia sakit. Bakteri dan virus memiliki kesamaan yaitu memiliki genom asam nukleat berdasarkan kode genetik yang sama digunakan dalam sel. Bakteri merupakan organisme partikel kecil sel satu dimana hidup dan berkembang biak tidak bergantung pada inang, lain halnya dengan virus yaitu dapat menular dengan mudah yang hidup dan berkembang biak melalui infeksi sel inang. Oleh karena itu infeksi bakteri dan virus dipandang sangat berbeda seperti antibiotik berfungsi untuk melawan bakteri. Jenis penyakit umumnya diakibatkan oleh virus salah satunya adalah flu biasa yang disebabkan oleh infeksi virus pada saluran pernapasan bagian atas. Sebagian virus umumnya hanya bersifat lokal namun virus COVID-19 berskala global dan merupakan jenis virus baru. Para ahli kesehatan berusaha memahami sifat virus COVID-19, namun hingga saat ini belum ada kelanjutan rincian bagaimana virus COVID-19 disebabkan dan langkah yang tepat untuk menangkal dan menghilangkan virus COVID-19. Para ahli kesehatan membuat vaksin tetapi tidak dapat meyakinkan semua orang di dunia dengan obat tersebut.

Novel coronavirus adalah salah satu virus berbahaya yang berasal dari Cina tepatnya provinsi Wuhan pada Desember 2019 sehingga menyebabkan pandemi global yang telah meregang banyak nyawa orang [1]. Virus COVID-19 merupakan kelompok virus “*Middle East Respiratory Syndrome CoronaVirus*” atau MERS-CoV yang mengakibatkan penyakit seperti batuk dan influenza, COVID-19 juga disebut “*Severe Acute Respiratory Syndrome CoronaVirus-2*” atau

SARS-CoV-2 [2]. Jumlah orang yang tertular virus ini meningkat secara signifikan dari hari ke hari. Jika hanya menggunakan masker, menjaga jarak satu meter dan mencuci tangan tidak dapat menurunkan peluang kita tertular virus atau menularkan ke orang lain. Dengan demikian dibutuhkan sebuah metode untuk menangkai penyebaran virus ini yaitu dengan vaksinasi. Vaksin berguna untuk memperkuat daya tahan tubuh kita untuk menyerang penyakit jika kita terpapar penyakit tersebut. Vaksinasi sudah diadakan di sebagian besar daerah yang terpapar virus COVID-19. Protokol vaksinasi yang aman, efisien, dan berskala besar adalah satu-satunya cara untuk menghilangkan pandemi yang sedang berlangsung yang telah menyebabkan masalah sosial dan ekonomi yang parah di seluruh dunia. Meskipun krisis pandemi sedang berlangsung, dunia tidak bisa menolak mengembangkan vaksin COVID-19 dalam jangka waktu pendek yang belum pernah terjadi sebelumnya [3]. Menurut keterangan WHO “vaksin COVID-19 untuk yang berusia 18 tahun ke atas dinyatakan aman, begitu juga yang sebelumnya mengalami masalah kesehatan dan gangguan autoimun, diabetes, asma, penyakit hati dan ginjal, serta infeksi kronis yang stabil dan terkontrol” [4].

Dari permasalahan diatas maka pada penelitian ini akan melakukan analisa ketercapaian vaksinasi di suatu wilayah tertentu di Indonesia. Vaksinasi diharapkan dapat memperkuat sistem kekebalan tubuh masyarakat dan kembali ke kehidupan normal. Untuk melihat perkembangan vaksinasi diperlukan alat yang dapat bekerja secara otomatis yang bekerja dari data tanpa pemrograman eksplisit yaitu menggunakan *Machine Learning*. *Machine Learning* adalah alat yang digunakan untuk penelitian yang membantu komputer dalam membuat pemodelan melalui pengalaman dan dapat memprediksi peristiwa di masa depan [5]. Pada *machine learning* terdapat kategori *Supervised learning* yaitu metode yang dilatih pada satu set data berlabel. Hal tersebut dapat dibagi menjadi 2 metode yaitu regresi dan klasifikasi. Teknik regresi digunakan untuk menggambarkan item data dengan prediksi aktual. Tujuan utama dari metode regresi adalah untuk menarik garis atau kurva yang paling cocok dari data.

Pada penelitian ini, data cakupan vaksinasi yang diprediksi adalah pada provinsi Jakarta dengan parameter yang akan di uji adalah sasaran, belum vaksin, dosis 1, dosis 2, total vaksin diberikan. Berdasarkan jenis data supervise learning maka model yang digunakan adalah *regressor* yaitu *AdaBoost*. Kinerja *regressor* ditentukan

berdasarkan *Mean Absolute Error (MAE)* dan *Root Mean Squared Error (RMSE)*.

1.2 Identifikasi Masalah

Berdasarkan latar belakang didapatkan beberapa rumusan masalah yang akan menjadi pembahasan pada penelitian ini yaitu:

1. Bagaimana melakukan analisa karakteristik dataset vaksinasi COVID-19?
2. Bagaimana menganalisa perkembangan vaksinasi COVID-19 di provinsi DKI Jakarta menggunakan *AdaBoost Regressor*?
3. Bagaimana mengetahui kinerja regressor menggunakan *MAE* dan *RMSE*?

1.3 Tujuan

Adapun tujuan dari penelitian ini adalah:

1. Mengetahui cara melakukan analisa karakteristik data pada dataset perkembangan vaksinasi COVID-19.
2. Mengetahui perkembangan vaksinasi di provinsi DKI Jakarta melalui algoritma *AdaBoost Regressor*.
3. Mengetahui kinerja regressor melalui *MAE* dan *RMSE*.

1.4 Manfaat

Adapun beberapa manfaat dari penelitian ini adalah:

1. Penelitian ini diharapkan bisa memberikan informasi, masukan dan referensi bagi pemerintah sebagai evaluasi dalam perkembangan vaksinasi COVID-19 di Provinsi Jakarta.
2. Bagi peneliti diharapkan penelitian ini dapat memberikan pengalaman belajar di masa pandemi COVID-19 sebagai dasar perbaikan diri dan meningkatkan kualitas belajar.

1.5 Ruang Lingkup

Dalam penelitian ini terdapat beberapa Batasan masalah agar penelitian ini tidak terlalu luas atau melebar yaitu:

1. Penelitian ini menggunakan dataset vaksinasi yang bersumber dari provinsi DKI Jakarta.
2. Parameter yang akan di uji pada penelitian ini adalah sasaran, belum vaksin, dosis 1, dosis 2, total vaksin diberikan.
3. Dataset yang digunakan dalam penelitian ini merupakan kumpulan data harian yang diambil pada 1 Juli 2021 – 30 November 2021.

Bab II

Landasan Teori

2.1 Penelitian Terdahulu

Pada penyusunan laporan penelitian ini membutuhkan rujukan yang diambil dari beberapa penelitian sebelumnya, yang tentunya rujukan tersebut berkaitan dengan topik penelitian yang dilakukan:

2.1.1 Penelitian Terkait

Tabel 2. 1 Penelitian terkait

No.	Area Penelitian	Karakteristik Data	Metode	Hasil Penelitian
1.	Memprediksi COVID-19 di negara Brazil [6].	Penelitian disini menggunakan data yang tersedia pada setiap negara bagian di Brazil dengan pembaruan angka harian.	<i>Machine Learning</i> .	Berdasarkan peluruhan kurva pada bulan September dan Oktober 2020 jumlah kasus aktif relative stabil, tetapi pada kuartal pertama tahun 2021 mengalami peningkatan. Temuan penting pada penerapan model SIRD ke Brazil menghasilkan prediksi yang berbeda untuk dinamika kasus, menunjukkan pengendalian epidemi pada Oktober 2020.

2.	Identifikasi pasien COVID-19 berisiko tinggi di Meksiko [7].	Data diambil dari data resmi yang disediakan oleh pemerintah Meksiko untuk umum, data berisi infeksi COVID-19 yang terkonfirmasi dan dicurigai di Meksiko.	<i>Machine Learning</i>	Hasil dari penelitian ini terbagi menjadi 4 golongan pasien. Pasien stadium 1 adalah pasien dalam perawatan medis awal. Pasien stadium 2 adalah pasien sebagai bagian dari evaluasi yang sudah memiliki riwayat COVID-19 dan sudah tertular pneumonia terkait COVID-19. Pasien stadium 3 adalah pasien yang sudah ditetapkan apakah dirawat atau tidak. Pasien stadium 4 adalah pasien yang sudah dirawat dan telah diinkubasi atau dirawat intensif.
3.	Prediksi dan Analisis COVID-19 di Wuhan, Cina [8].	Data diambil berdasarkan data dari Wuhan dari tanggal 23 Januari 2020 hingga 6 April 2020, dengan variabel kasus terkonfirmasi kumulatif, kematian kumulatif, dan kasus sembuh kumulatif.	<i>Machine Learning</i>	Hasil dari penelitian tersebut adalah dari tanggal 23 Januari 2020 sampai dengan 11 Februari 2020, jumlah kasus baru terkonfirmasi menunjukkan tren kenaikan yang fluktuatif. kasus baru yang dikonfirmasi berfluktuasi tajam pada 12 Februari, menunjukkan puncaknya, alasannya

				<p>adalah pengumuman 12.364 kasus yang dikonfirmasi secara klinis untuk pertama kalinya pada hari itu, dengan peningkatan tajam dalam kasus baru yang dikonfirmasi. Jumlah kasus baru sembuh menunjukkan tren meningkat sebelum 27 Februari, setelah itu tren pertumbuhan kasus sembuh melambat. Selain itu, jumlah kematian baru menunjukkan tren penurunan yang signifikan setelah 24 Februari, menunjukkan pemulihan yang menjanjikan di masa depan.</p>
4.	<p>Memprediksi COVID-19 uji klinis selesai vs penghentian [9].</p>	<p>Data berjumlah 4.441 yang bersal dari ClinicalTrial.gov dengan variabel uji klinis, kelayakan, informasi studi, kriteria, jenis obat, kata kunci</p>	<p><i>Machine Learning</i></p>	<p>Hasil penelitian menunjukkan bahwa fitur obat dan kata kunci studi adalah fitur yang paling informatif, tetapi keempat jenis fitur sangat penting untuk prediksi percobaan yang akurat. Berdasarkan hasil penelitian menunjukkan bahwa</p>

		studi		metode komputasi dapat memberikan fitur yang efektif untuk memahami perbedaan antara selesai vs menghentikan uji coba COVID-19. Selain itu, model tersebut juga dapat memprediksi status uji coba COVID-19 dengan akurasi yang memuaskan dengan skor 0,87 AUC dan akurasi 0,81.
5.	Memprediksi hasil buruk pasien covid-19 pada hari masuk di New York City, Amerika Serikat [10].	Data berasal dari Pusat Medis Irving Universitas Columbia dari 10 Maret 2020, hingga 24 Mei 2020	<i>Machine Learning</i>	Dari 2.545 pasien, 833 (32,7%) mengalami titik akhir primer. 53 tes laboratorium dianalisis, dan di antaranya, 47 tes (dan usia) berbeda secara signifikan antara pasien dengan dan tanpa titik akhir. Model multivariat terakhir termasuk usia, albumin, kreatinin, protein C-reaktif, dan laktat dehidrogenase. Area di bawah kurva ROC adalah 0,850 (CI [95%]: 0,813, 0,889), dengan sensitivitas 0,800 dan spesifisitas 0,761.
6.	Tren	Data berasal	<i>Machine</i>	Hasil eksperimen

	penularan Covid-19 dan penyebaran wabah virus di India [11].	dari dataset kehidupan nyata dari Kementerian Kesehatan dan Kesejahteraan Keluarga India.	<i>Learning</i>	menunjukkan bahwa Regresi Polinomial menunjukkan akurasi yang lebih baik (93% appx) daripada SVM untuk memprediksi peningkatan kasus dalam 60 hari ke depan. Sebagai kesimpulan, disebutkan bahwa ke depan model deep learning atau model hybrid dapat digunakan untuk meramalkan penyebaran virus
7.	Memprediksi hasil tes covid-19 dengan akurasi tinggi di Israel [12].	Dataset berasal dari Kementerian Kesehatan Israel dengan variabel yang digunakan jenis kelamin, usia 60 tahun, diketahui kontak dengan individu yang terinfeksi, dan munculnya lima gejala klinis awal.	<i>Machine Learning</i>	Model diprediksi dengan 0,09 auROC (area di bawah kurva karakteristik operasi penerima) dengan 95% CI: 0,892-0,905 dan dengan menggunakan prediksi dari set uji, titik kerja yang mungkin adalah: sensitivitas 87,30% dan spesifisitas 71,98% atau sensitivitas 85,76% dan spesifisitas 79,18%
8.	Survei global tentang potensi penerimaan vaksin	Data dari 16 Juni hingga 20 Juni 2020, 13426 orang dari 19 negara	<i>Machine Learning</i>	Hasilnya, mereka melihat tingkat tertinggi di China (hampir 90%) dan tingkat terendah di

	COVID-19 [13].	untuk menentukan tingkat penerimaan Vaksin COVID-19		Rusia (hampir 55%). Dari total dataset, 71,5% menjawab bahwa mereka akan mengambil vaksin jika terbukti aman. dan 48% menjawab bahwa mereka akan mengambil Vaksin jika majikan mereka merekomendasikannya
9.	Dampak vaksinasi menurut kelompok prioritas pada kematian di Inggris [14].	Data diambil dari Departemen Kesehatan dan Perawatan Sosial. Dengan variabel karakteristik individu yang disertakan (usia dalam tahun), Jumlah total (jutaan), Proporsi populasi dewasa Inggris Raya, Proporsi kumulatif populasi dewasa Inggris Raya, Proporsi kumulatif	<i>Machine Learning</i>	Berdasarkan hasil penelitian menemukan bahwa vaksinasi menurunkan kematian akibat COVID-19, rawat inap, dan penerimaan di ICU, dengan dampak yang lebih cepat pada kematian akibat COVID-19. Para penulis juga menunjukkan bahwa dampak vaksinasi pada rawat inap dan penerimaan ICU dapat lebih ditingkatkan dengan memprioritaskan awal staf kesehatan dan pasien yang rentan secara klinis.

		kelompok yang disertakan		
10.	Kemajuan vaksinasi di seluruh dunia [15].	Data diambil dari repositori Kaggle dengan rentang waktu 14 Desember 2020 dan 24 April 2021.	<i>Machine Learning</i>	Visualisasi laju vaksinasi aktual dan prediksi yang diperoleh dengan model Regressor yaitu Random Forest Regressor, Extra Trees Regressor, Gradient Boosting Regressor, AdaBoostRegressor dan XGB Regressor. AdaBoostRegressor mengungguli dengan meminimalkan kesalahan absolut rata-rata (MAE) sebesar 9,968 dan kesalahan kuadrat rata-rata akar (RMSE) sebesar 11,133.

2.2 Landasan Teori

2.2.1 Vaksinasi

Pandemi COVID-19 sedang berlangsung diseluruh dunia. COVID-19 merupakan virus berbahaya yang dapat menular dan sudah merenggut banyak nyawa orang maka dibutuhkan salah satu alat kesehatan yang dapat melindungi imun tubuh untuk mengurangi penyebaran dan bahaya yang disebabkan yaitu vaksinasi [16]. Sebelum COVID-19 ada, vaksinasi sudah ada untuk penyakit lainnya. Berdasarkan data *World Health Organization (WHO)* pada tahun 2010 – 2015 vaksin sudah mencegah 10 juta kematian [16].

2.2.2 Perkembangan Vaksinasi

Jumlah orang yang tertular virus ini meningkat secara signifikan dari hari ke hari. Jika hanya menggunakan masker, menjaga jarak satu meter dan mencuci tangan tidak dapat menurunkan peluang kita tertular virus atau menularkan ke orang lain. Dengan demikian dibutuhkan sebuah metode untuk menangkal penyebaran virus ini yaitu dengan vaksinasi. Vaksin berguna untuk memperkuat daya tahan tubuh kita untuk menyerang penyakit jika kita terpapar penyakit tersebut. Vaksinasi sudah diadakan di sebagian besar daerah yang terpapar virus COVID-19. Protokol vaksinasi yang aman, efisien, dan berskala besar adalah satu-satunya cara untuk menghilangkan pandemi yang sedang berlangsung yang telah menyebabkan masalah sosial dan ekonomi yang parah di seluruh dunia. Meskipun krisis pandemi sedang berlangsung, dunia tidak bisa menolak mengembangkan vaksin COVID-19 dalam jangka waktu pendek yang belum pernah terjadi sebelumnya.

Maka daripada itu adapun tujuan penelitian ini akan melakukan analisa ketercapaian vaksinasi di suatu wilayah tertentu di Indonesia. Vaksinasi diharapkan dapat memperkuat sistem kekebalan tubuh masyarakat dan kembali ke kehidupan normal. Untuk melihat perkembangan vaksinasi diperlukan alat yang dapat bekerja secara otomatis yang bekerja dari data tanpa pemrograman eksplisit yaitu menggunakan Machine Learning. Machine Learning adalah alat yang digunakan untuk penelitian yang membantu komputer dalam membuat pemodelan melalui pengalaman dan dapat memprediksi peristiwa di masa depan

2.2.3 Regresi

Regresi merupakan suatu model pada supervised learning atau pembelajaran yang diawasi untuk permasalahan dalam analisis regresi di mana variabel yang menarik mengasumsikan nilai dalam interval biner [17]. Regresi merupakan suatu model pembelajaran mesin yang paling penting untuk prediksi data deret waktu. Namun, multikolinearitas juga menjadi kendala akurasi. Pendekatan yang paling mudah (disebut Tipe 1) untuk mengatasi masalah kolinearitas adalah dengan menghilangkan beberapa variabel terkait sebelum pemodela [18]. Regresi mengasumsikan bahwa data target cocok dengan beberapa tipe fungsi yang diketahui dari tipe ini yang memodelkan kumpulan data yang diberikan.

2.2.4 AdaBoost or Adaptive Boosting Regressor

Algoritma Adaboost adalah kerangka belajar ensemble yang pelajar lemah dari algoritma. Adaboost dapat dengan mudah diganti untuk menemukan solusi terbaik untuk masalah yang sesuai [19]. AdaBoost merupakan salah satu algoritma supervised learning yang berkonsentrasi untuk mengamati pada klasifikasi yang salah dan meningkatkan bobot pengamatan tersebut dan mengubah distribusi data secara adaptif. Ini mendefinisikan properti pemasangan data pelatihan sepenuhnya tanpa kesalahan [20]. AdaBoost Regressor dapat menghasilkan regresi berganda untuk menyelesaikan model pembobotan dan secara otomatis menyesuaikan pembobotan berdasarkan kesalahan estimasi. Oleh karena itu, ia memiliki potensi besar untuk memecahkan masalah regresi yang kompleks dan non-linier. AdaBoost Regressor menunjukkan bahwa banyak proses degradasi memiliki karakteristik non-linier yang kompleks [21]. Adapun rumus pada Adaboost Regressor adalah:

$$G_m(x) = \operatorname{argmax}[\sum_{m=1}^M \alpha_m G_m(x)]$$

2.2.5 Regressor Model Evaluation dan Mean Absolute Error

Dua ukuran *absolute error* yang banyak digunakan dalam disiplin ilmu machine learning dan analisis time series adalah *Root Mean Square Error* (RMSE) dan *Mean Absolute Error* (MAE). MAE dan RMSE merupakan metrik yang memperkirakan kuantitas yang sama dalam faktor skala [22]. Prediksi sempurna diperoleh ketika nilai MAE adalah 0. RMSE adalah akar kuadrat dari Mean Squared Error (MSE). MSE adalah perbedaan kuadrat antara nilai aktual dan prediksi. Semakin rendah nilai RMSE, semakin tinggi efisiensi model [15].

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2}$$

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |e_t|$$

e_t = nilai aktual

n = banyaknya data

Σ = Jumlah keseluruhan nilai

2.2.6 *Python*

Python adalah bahasa pemrograman yang secara umum bisa digunakan untuk beberapa bidang seperti data statistik, web, database. Berbeda dengan *R* yang hanya berfokus untuk bidang komputasi statistik. *Python* memiliki dua versi yaitu versi 2.x dan versi 3.x, yang mana keduanya tidak sepenuhnya kompatibel meskipun adanya kemiripan. Versi 2.x adalah versi lama yang pemeliharaannya sudah berakhir pada tahun 2020. Versi 3.x merupakan desain ulang versi sebelumnya yang dianggap sebagai masa depan *python* [23].

2.2.7 *Matplotlib*

Matplotlib adalah alat analisis yg disediakan untuk *python* yang sangat lengkap untuk membuat visualisasi, animasi, dan interaktif. *Matplotlib* dapat membuat hal mudah menjadi mudah dan hal sulit menjadi mungkin [24]. Adapun cara pemasangan *matplotlib* seperti berikut:

- pip install Matplotlib

2.2.8 *Seaborn*

Seaborn adalah alat analisis yang disediakan untuk *python* yang menyediakan antarmuka tingkat tinggi dan menggambar grafik statistik yang menarik dan informatif berdasarkan *matplotlib* [25]. Adapun cara pemasangan *seaborn* seperti berikut:

- pip install seaborn

2.2.9 *Pandas*

Pandas adalah alat analisis dan manipulasi data sumber terbuka yang cepat, kuat, fleksibel, dan mudah digunakan, dibangun menggunakan bahasa pemrograman *Python* [26]. Adapun cara pemasangan *pandas* seperti berikut:

- pip install pandas

2.2.10 *Numpy*

Numpy adalah alat analisis yang *open source* yang dibuat pada tahun 2005 yang bertujuan untuk komputasi numerik pada python. NumPy akan selalu menjadi perangkat lunak *open source* 100%, gratis untuk digunakan semua orang dan dirilis di bawah persyaratan liberal dari lisensi BSD yang dimodifikasi [27]. Adapun cara pemasangan pandas seperti berikut:

- `pip install numpy`

2.2.11 *Scikit-learn*

Sklearn atau *scikit-learn* adalah suatu alat analisis yang simpel dan efisien yang digunakan untuk analisis data prediktif yang dapat digunakan secara komersial. *Sklearn* sering digunakan untuk *machine learning* yang tersedia pada *python*. *Sklearn* dibangun berdasarkan *Numpy*, *Scipy*, dan *matplotlib*. *Sklearn* juga salah satu library *python* yang bersifat *open source* yang dirilis dan dapat digunakan kembali dan dirilis melalui lisensi BSD [28]. Adapun cara pemasangan Scikit-learn seperti berikut :

- `pip install -U scikit-learn`

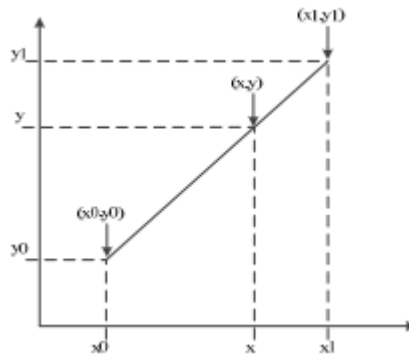
2.2.12 *Outlier*

Secara teoritis outlier adalah objek yang menyimpang dari objek lainnya yang dapat menimbulkan keraguan terhadap objek tersebut dihasilkan oleh mekanisme yang berbeda. Secara teoritis terdapat berbagai definisi outlier berdasarkan karakteristik pengamatan yang berbeda dari objek data [29]. Sebuah objek dikatakan outlier jika :

- Menyimpang dari perilaku data yang normal,
- Mengasumsikan nilai yang jauh dari nilai yang diharapkan/rata-rata,
- Tidak berhubungan/mirip dengan objek lain dalam hal karakteristiknya.

2.2.13 Interpolasi Linier

Interpolasi linier adalah suatu metode interpolasi untuk suatu dimensi pada data. Dengan memperkirakan nilai data berdasarkan dua titik data yang berdekatan dengan titik yang perlu diinterpolasi dalam urutan data satu dimensi [30]. Sebagai contoh salah satunya adalah interpolasi linier yang paling sederhana yaitu interpolasi linier orde satu seperti berikut :



Gambar 2. 1 Grafik interpolasi linier

Adapun rumus yang dapat digunakan untuk perhitungan menggunakan interpolasi linier adalah:

$$y(x) = y_1 + (x - x_1) \frac{(y_2 - y_1)}{(x_2 - x_1)}$$

x_1, y_1 adalah koordinat titik pertama

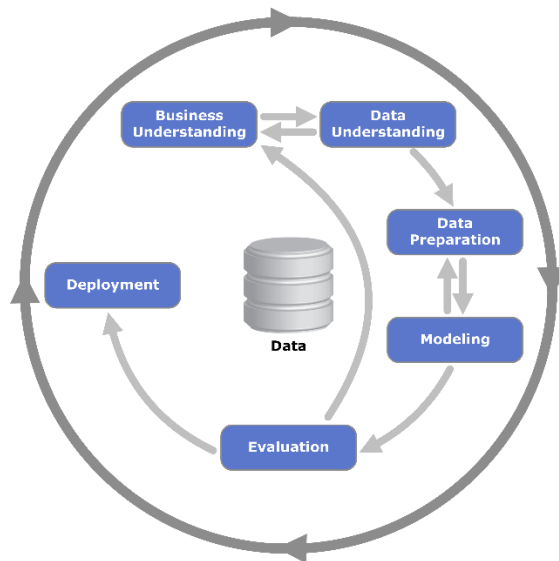
x_2, y_2 adalah koordinat titik kedua

x adalah titik dimana dilakukan interpolasi

y adalah nilai interpolasi

2.3 Metodologi Penelitian

Metodologi penelitian berasal dari kata “metode” yang artinya cara yang tepat untuk melakukan sesuatu; dan “logos” yang artinya ilmu atau pengetahuan. Jadi secara harfiah metodologi penelitian adalah cara melakukan sesuatu dengan menggunakan pikiran secara saksama untuk mencapai suatu



Gambar 2. 2 Alur diagram CRISP-DM

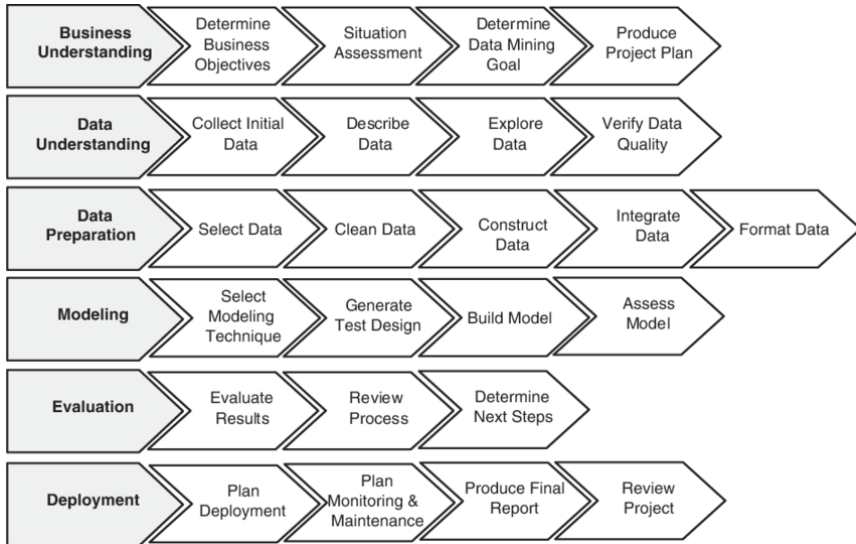
Penjelasan dari alur metodologi penelitian pada gambar diatas adalah sebagai berikut:

- a. *Bussines Understanding* : Pada tahap ini bertujuan pada pemahaman proyek dan persyaratan dalam bisnis lalu mengubah pengetahuan tersebut menjadi definisi masalah penambahan data dan planning yang dirancang untuk mencapai tujuan.
- b. *Data Understanding* : Pada tahap ini dilakukan pengumpulan data dan mengidentifikasi masalah data untuk mendeteksi hambatan yang menarik untuk informasi.
- c. *Data Preparation* : Pada tahap ini mencakup semua kegiatan untuk membangun dataset akhir dari data mentah. Tahap ini dapat dilakukan beberapa kali dan tidak ada ketentuan

urutannya. Pada tahap ini dilakukan pemilihan tabel, membersihkan data, transformasi data untuk masuk pemodelan.

- d. *Modelling* : Pada tahap ini dilakukan pemilihan Teknik pemodelan kemudian diterapkan dan parameternya dikalibrasikan ke nilai optimal. Beberapa teknik memilih persyaratan spesifik pada bentuk data. Oleh karena itu mundur ke fase *data preparation* sering dilakukan.
- e. *Evaluation* : Pada tahap ini telah dibangun model yang tampaknya memiliki kualitas tinggi melalui perspektif analisa data. Sebelum memasuki tahap deployment tahap evaluasi sangat penting untuk mencapai tujuan bisnis.
- f. *Deployment* : Penciptaan model bukan akhir dari proyek. Sementara tujuan dari model adalah untuk meningkatkan wawasan dari data, wawasan yang diperoleh harus diatur dan disajikan dengan cara yang dapat digunakan oleh pelanggan. Ini sering melibatkan penerapan model "LIVE" dalam proses pengambilan keputusan organisasi, seperti personalisasi, halaman web waktu nyata, atau evaluasi ulang berulang dari basis data pemasaran, sesederhana menghasilkan laporan atau seruit menerapkan laporan berulang proses penambangan data di seluruh perusahaan.

Namun pada alur metodologi juga terdapat langkah yang lainnya seperti gambar berikut:



Gambar 2. 3 Langkah Diagram CRISP-DM

Pada Business Understanding terdapat beberapa Langkah sebagai berikut:

- Determine business objectives atau menentukan tujuan bisnis. Tujuan pertama dari analis data adalah untuk benar-benar memahami, dari perspektif bisnis, apa yang benar-benar ingin dicapai oleh klien. Seringkali klien memiliki banyak tujuan dan kendala yang saling bersaing yang harus diseimbangkan dengan baik. Tujuan analis adalah untuk mengungkap faktor-faktor penting yang dapat mempengaruhi hasil proyek. Kemungkinan konsekuensi dari mengabaikan langkah ini adalah mengeluarkan banyak usaha untuk menghasilkan jawaban yang benar atas pertanyaan yang salah.
- Assess situation atau menilai situasi. Tugas ini melibatkan pencarian fakta yang lebih rinci tentang semua sumber daya, kendala, asumsi, dan faktor lain yang harus dipertimbangkan dalam menentukan tujuan analisis data dan rencana proyek.

- Determine data mining goals atau menentukan tujuan penambangan data. Tujuan bisnis menyatakan tujuan dalam terminologi bisnis. Tujuan penambangan data menyatakan tujuan proyek dalam istilah teknis. Misalnya, sasaran bisnis mungkin "Meningkatkan penjualan katalog ke pelanggan yang sudah ada". Tujuan penambangan data mungkin adalah "Memprediksi berapa banyak widget yang akan dibeli pelanggan, mengingat pembelian mereka selama tiga tahun terakhir, informasi demografis (usia, gaji, kota, dll.) dan harga item".
- Produce project plan atau menghasilkan rencana proyek. Jelaskan rencana yang dimaksudkan untuk mencapai tujuan penambangan data dan dengan demikian mencapai tujuan bisnis. Rencana tersebut harus merinci serangkaian langkah antisipasi yang akan dilakukan selama sisa proyek termasuk pemilihan alat dan teknik awal.

Pada Data Understanding terdapat beberapa Langkah sebagai berikut:

- Collect initial data atau mengumpulkan data awal. Pengumpulan awal ini mencakup pemuatan data jika diperlukan untuk pemahaman data. Misalnya, jika Anda menerapkan alat khusus untuk pemahaman data, masuk akal untuk memuat data Anda ke alat ini.
- Describe data atau menjelaskan data. Periksa properti "kotor" atau "permukaan" dari data yang diperoleh dan laporkan hasilnya.
- Explore data atau menjelajahi data. Tugas ini menangani pertanyaan penambangan data, yang dapat diatasi menggunakan kueri, visualisasi, dan pelaporan.
- Verify data quality atau memverifikasi kualitas data. Periksa kualitas data, jawab pertanyaan seperti: apakah datanya lengkap (apakah mencakup semua kasus yang diperlukan)? Apakah benar atau mengandung kesalahan dan jika ada

kesalahan seberapa umumkah kesalahan tersebut? Apakah ada nilai yang hilang dalam data?

Pada Data Preparation terdapat beberapa Langkah sebagai berikut:

- Select data atau memilih data. Tentukan data yang akan digunakan untuk analisis. Kriteria termasuk relevansi dengan tujuan penambangan data, kualitas dan kendala teknis seperti batasan volume data atau tipe data.
- Clean data atau membersihkan data. Tingkatkan kualitas data ke tingkat yang dibutuhkan oleh teknik analisis yang dipilih.
- Construct data atau membangun data. Tugas ini mencakup operasi persiapan data konstruktif seperti produksi atribut turunan, seluruh catatan baru, atau nilai yang diubah untuk atribut yang ada.
- Integrate data atau mengintegrasikan data. Ini adalah metode dimana informasi digabungkan dari beberapa tabel atau record untuk membuat record atau nilai baru.
- Format data. Transformasi pemformatan mengacu pada modifikasi sintaksis yang dibuat pada data yang tidak mengubah artinya, tetapi mungkin diperlukan oleh alat pemodelan.

Pada modelling terdapat beberapa Langkah sebagai berikut:

- Select modeling technique atau memilih teknik pemodelan. Sebagai langkah pertama dalam pemodelan, pilih teknik pemodelan aktual yang akan digunakan. Meskipun Anda mungkin sudah memilih alat dalam pemahaman bisnis, tugas ini mengacu pada teknik pemodelan khusus.
- Generate test design atau menghasilkan desain pengujian. Sebelum kita benar-benar membangun model, kita perlu membuat prosedur atau mekanisme untuk menguji kualitas dan validitas model. Misalnya, dalam tugas penambangan data yang diawasi seperti klasifikasi, biasanya menggunakan

tingkat kesalahan sebagai ukuran kualitas untuk model penambangan data. Oleh karena itu, kami biasanya memisahkan dataset menjadi train dan test set, membangun model pada train set dan memperkirakan kualitasnya pada test set yang terpisah.

- Build model atau membangun model. Jalankan alat pemodelan pada kumpulan data yang disiapkan untuk membuat satu atau beberapa model.
- Assess model atau menilai model. Insinyur data mining menginterpretasikan model sesuai dengan pengetahuan domainnya, kriteria keberhasilan data mining dan desain pengujian yang diinginkan.

Pada evaluation terdapat beberapa Langkah sebagai berikut:

- Evaluate results atau evaluasi hasil. Langkah-langkah evaluasi sebelumnya berkaitan dengan faktor-faktor seperti akurasi dan keumuman model. Langkah ini menilai sejauh mana model memenuhi tujuan bisnis dan berusaha untuk menentukan apakah ada beberapa alasan bisnis mengapa model ini kurang. Pilihan evaluasi lainnya adalah menguji model pada aplikasi uji dalam aplikasi nyata jika kendala waktu dan anggaran memungkinkan.
- Review process atau proses peninjauan. Pada titik ini model yang dihasilkan diharapkan tampak memuaskan dan memenuhi kebutuhan bisnis. Sekarang tepat untuk melakukan tinjauan yang lebih menyeluruh terhadap keterlibatan penambangan data untuk menentukan apakah ada faktor atau tugas penting yang entah bagaimana telah diabaikan.
- Determine next steps atau menentukan langkah selanjutnya. Menurut hasil penilaian dan tinjauan proses, proyek memutuskan bagaimana melanjutkan pada tahap ini. Proyek perlu memutuskan apakah akan menyelesaikan proyek ini dan melanjutkan ke penerapan jika sesuai atau apakah akan

memulai iterasi lebih lanjut atau menyiapkan proyek penambangan data baru.

Pada deployment terdapat beberapa Langkah sebagai berikut:

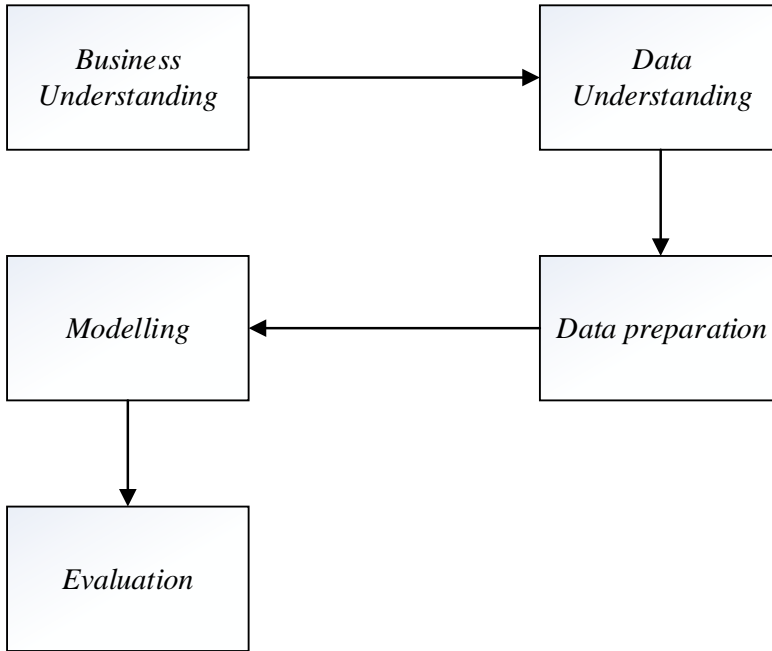
- Plan deployment atau merencanakan penerapan. Untuk menyebarkan hasil penambangan data ke dalam bisnis, tugas ini mengambil hasil evaluasi dan menyimpulkan strategi untuk penyebaran. Jika prosedur umum telah diidentifikasi untuk membuat model yang relevan, prosedur ini didokumentasikan di sini untuk penerapan selanjutnya.
- Plan monitoring and maintenance atau merencanakan pemantauan dan pemeliharaan. Pemantauan dan pemeliharaan merupakan isu penting jika hasil data mining menjadi bagian dari bisnis sehari-hari dan lingkungannya. Persiapan strategi pemeliharaan yang cermat membantu menghindari penggunaan hasil data mining yang salah dalam jangka waktu lama yang tidak perlu.
- Produce final report atau menghasilkan laporan akhir. Di akhir proyek, pemimpin proyek dan timnya menulis laporan akhir. Bergantung pada rencana penerapan, laporan ini mungkin hanya ringkasan proyek dan pengalamannya.
- Review project atau meninjau proyek. Menilai apa yang berjalan dengan benar dan apa yang salah, apa yang dilakukan dengan baik dan apa yang perlu ditingkatkan.

Bab III

Metodologi Penelitian

3.1 Metode Penelitian

Pada penelitian kali ini penulis menggunakan metode penelitian dengan mengikuti tahapan model *Cross-Industry Standard Process for Data Mining (CRISP-DM)*. Adapun tahapan tahapan pada *CRISP-DM* yaitu *Business Understanding, Data Understanding, Data preparation, Modelling, Evaluation, Deployment*. Dalam penelitian kali ini, penulis membuat beberapa tahapan sebagai berikut:



Gambar 3. 1 Alur Diagram Metodologi Penelitian

3.2 Tahapan-tahapan Diagram Alur

Tahapan metodologi penelitian diatas dapat diuraikan dan dijelaskan lebih detail seperti berikut:

3.2.1 Business Understanding

Tujuan	: Pahaman tujuan proyek
Masukan	: -
Metode	: Studi literatur
Keluaran	: Mendefinisikan masalah <i>data mining</i> dan membuat <i>planning</i> untuk mencapai tujuan

Pada tahap ini memahami tujuan proyek pada perspektif bisnis kemudian mengubah pemahan tersebut untuk mendefinisikan masalah pada *data mining* dan membuat *planning* untuk mencapai tujuan sehingga proses berjalannya proyek menjadi lancar tanpa permasalahan.

3.2.2 Data Understanding

Tujuan : Mengumpulkan data

Masukan : Data sekunder

Metode : Studi literatur

Keluaran : Data berbentuk excel

Pada tahap ini melakukan proses pengumpulan data, kemudian melakukan analisa terhadap data untuk dapat membangun data dan membuat model terbaik untuk data. Pada tahap ini penulis mengumpulkan data melalui website satgas COVID-19 DKI Jakarta dari tanggal 1 Juli hingga 30 November 2021. Data yang dikumpulkan adalah data sekunder. Data sekunder adalah data yang sudah diolah sebelumnya. Sumber data sekunder bisa berupa jurnal atau publikasi pemerintah.

3.2.3 Data preparation

Tujuan : Memilih data dan membersihkan data

Masukan : Data sekunder

Metode : Studi literatur

Keluaran : Ringkasan data

Pada tahap ini dilakukan peninjauan data, memilih data dan membersihkan data saat menemukan masalah pada saat pembangunan model. Tahap ini dilakukan untuk memberikan fondasi analitik dengan membuat ringkasan data. Tahap ini tidak boleh dilakukan terburu buru seperti visualisasi data, jika ada masalah pada tahap ini maka akan mengganggu *modelling* nantinya. Pada tahap ini juga dilakukan *sampling* yaitu membagi daya menjadi

2 yaitu *data training* dan *data testing* sehingga memudahkan proses pada tahap pemodelan.

3.2.4 Modelling

Tujuan	: Mengolah data menggunakan model <i>AdaBoost Regressor</i>
Masukan	: Data sekunder
Metode	: Studi literatur
Keluaran	: Hasil

Dalam tahap ini metode statistika dan *machine learning* untuk penentuan terhadap *data mining* lalu pemodelan dipilih dan diterapkan ke dataset yang sudah disiapkan untuk mengatasi kebutuhan bisnis yang sesuai. Metode yang digunakan pada penelitian kali ini yaitu *AdaBoost Regressor*. Lalu kemudian nantinya keluar hasil dari pemodelan yang dilakukan. Jika diperlukan penyesuaian data kembali pada tahap ini maka dapat kembali pada tahap *preparation*.

3.2.5 Evaluation

Tujuan	: Menguji dan mengevaluasi model
Masukan	: Data sekunder
Metode	: Studi literatur
Keluaran	: Kesimpulan

Pada tahap ini, model yang sudah dibuat diuji dan dievaluasi keakuratan. Tahap ini mengukur sejauh mana model yang sudah dipilih memenuhi sasaran-sasaran bisnis dan bila demikian, sejauh manakah itu (apakah perlu lebih banyak model untuk dibuat dan diukur).

Bab IV

Analisis dan Hasil Pembahasan

4.1 Analisis

Pada tahap analisis dan perancangan sistem yang akan dibangun arahan ini akan digunakan sebagai arahan dalam membangun sistem agar sesuai dengan kebutuhan yang dibutuhkan serta mencegah hal-hal yang tidak dibutuhkan sehingga membuat permasalahan yang ada.

4.1.1 Data yang digunakan

Pada penelitian kali ini digunakan data set yang dikumpulkan melalui website satgas COVID-19 DKI Jakarta dari tanggal 1 Juli sampai 30 November tahun 2021. Data yang digunakan yaitu data perkelurahan di provinsi DKI Jakarta yaitu 267 kelurahan 44 kecamatan 5 Kotamadya dan 1 Kabupaten. Data yang digunakan bersifat kumulatif yang dikumpulkan perhari. Attribut yang digunakan yaitu “SASARAN” yaitu sasaran atau target berapa banyak penduduk yang akan di vaksin nantinya. Kemudian attribute selanjutnya yaitu “BELUM VAKSIN” yaitu berapa sisa warga yang belum di vaksin berdasarkan target atau sasaran. Lalu attribute digunakan yaitu “DOSIS 1” yaitu jumlah vaksin dosis 1 yang sudah diberikan kepada penduduk, “DOSIS 2” yaitu jumlah vaksin dosis 2 yang sudah diberikan kepada penduduk, “TOTAL” yaitu penjumlahan dosis vaksin 1 dan vaksin dosis 2 yang sudah di berikan pada penduduk.

Data perharinya terdapat 267 baris kemudian digabungkan sehingga total data menjadi 37380 baris. Berikut data yang digunakan dalam penelitian.

IV-29 | Analisis dan Hasil Pembahasan

TANGGAL	KOTA	KECAMATAN	KELURAHAN	SASARAN	BELUM VAKSIN	DOSIS1	DOSIS2	TOTAL
2021-07-01	JAKARTA UTARA	PADEMANGAN	ANCOL	20393	13272	7114	3287	10401
2021-07-01	JAKARTA BARAT	TAMBORA	ANGKE	25785	16477	9299	3221	12520
2021-07-01	JAKARTA TIMUR	KRAMAT JATI	BALE KAMBANG	25158	18849	6301	2644	8945
2021-07-01	JAKARTA TIMUR	JATINEGARA	BALI MESTER	8683	5743	2937	1517	4454
2021-07-01	JAKARTA TIMUR	CIPAYUNG	BAMBU APUS	22768	15407	7357	3985	11342
...
2021-11-30	JAKARTA TIMUR	MATRAMAN	UTAN KAYU SELATAN	33511	5753	27758	23688	51446
2021-11-30	JAKARTA TIMUR	MATRAMAN	UTAN KAYU UTARA	28169	5448	22721	19074	41795
2021-11-30	JAKARTA PUSAT	KEMAYORAN	UTAN PANJANG	29208	5515	23693	19731	43424
2021-11-30	JAKARTA UTARA	TANJUNG PRIOK	WARAKAS	44401	7828	36573	30403	66976
2021-11-30	JAKARTA BARAT	GROGOL PETAMBURAN	WIJAYA KUSUMA	38934	7741	31193	27022	58215

37380 rows x 8 columns

Gambar 4. 1 Tampilan data

TANGGAL	KOTA	KECAMATAN	KELURAHAN	SASARAN	BELUM VAKSIN	DOSIS1	DOSIS2	TOTAL
2021-07-01	JAKARTA UTARA	PADEMANGAN	ANCOL	20393	13272	7114	3287	10401
2021-07-01	JAKARTA BARAT	TAMBORA	ANGKE	25785	16477	9299	3221	12520
2021-07-01	JAKARTA TIMUR	KRAMAT JATI	BALE KAMBANG	25158	18849	6301	2644	8945
2021-07-01	JAKARTA TIMUR	JATINEGARA	BALI MESTER	8683	5743	2937	1517	4454
2021-07-01	JAKARTA TIMUR	CIPAYUNG	BAMBU APUS	22768	15407	7357	3985	11342
2021-07-01	JAKARTA SELATAN	MAMPANG PRAPATAN	BANGKA	18930	12503	6421	2704	9125
2021-07-01	JAKARTA TIMUR	PASAR REBO	BARU	20267	11268	8982	4674	13656
2021-07-01	JAKARTA TIMUR	KRAMAT JATI	BATU AMPAR	41389	30358	11020	5254	16274
2021-07-01	JAKARTA PUSAT	TANAH ABANG	BENDUNGAN HILIR	19008	11502	7499	3566	11065
2021-07-01	JAKARTA TIMUR	JATINEGARA	BIDARA CINA	32331	23395	8932	4034	12966
2021-07-01	JAKARTA SELATAN	PESANGGRAHAN	BINTARO	45843	31042	14790	7070	21860
2021-07-01	JAKARTA SELATAN	TEBET	BUKIT DURI	29934	20254	9678	3957	13635
2021-07-01	JAKARTA PUSAT	SENEN	BUNGUR	15932	9968	5948	2739	8687
2021-07-01	JAKARTA TIMUR	CAKUNG	CAKUNG BARAT	49369	32912	16454	4011	20465
2021-07-01	JAKARTA TIMUR	CAKUNG	CAKUNG TIMUR	50344	33061	17273	6385	23658

Gambar 4. 2 Tampilan Data Head

TANGGAL	KOTA	KECAMATAN	KELURAHAN	SASARAN	BELUM VAKSIN	DOSIS1	DOSIS2	TOTAL
2021-11-30	JAKARTA UTARA	TANJUNG PRIOK	TANJUNG PRIOK	34588	6847	27741	23273	51014
2021-11-30	JAKARTA SELATAN	TEBET	TEBET BARAT	21194	3256	17938	16175	34113
2021-11-30	JAKARTA SELATAN	TEBET	TEBET TIMUR	18199	2799	15400	14243	29643
2021-11-30	JAKARTA BARAT	KALI DERES	TEGAL ALUR	83647	16560	67087	56946	124033
2021-11-30	JAKARTA SELATAN	MAMPANG PRAPATAN	TEGAL PARANG	32401	6842	25559	20743	46302
2021-11-30	JAKARTA BARAT	GROGOL PETAMBURAN	TOMANG	29605	5233	24372	21558	45930
2021-11-30	JAKARTA UTARA	KOJA	TUGU SELATAN	39122	8367	30755	25706	56461
2021-11-30	JAKARTA UTARA	KOJA	TUGU UTARA	67670	15359	52311	42769	95080
2021-11-30	JAKARTA TIMUR	CAKUNG	UJUNG MENTENG	29934	4938	24996	21836	46832
2021-11-30	JAKARTA SELATAN	PESANGGRAHAN	ULUJAMI	40544	8647	31897	27107	59004
2021-11-30	JAKARTA TIMUR	MATRAMAN	UTAN KAYU SELATAN	33511	5753	27758	23688	51446
2021-11-30	JAKARTA TIMUR	MATRAMAN	UTAN KAYU UTARA	28169	5448	22721	19074	41795
2021-11-30	JAKARTA PUSAT	KEMAYORAN	UTAN PANJANG	29208	5515	23693	19731	43424
2021-11-30	JAKARTA UTARA	TANJUNG PRIOK	WARAKAS	44401	7828	36573	30403	66976

Gambar 4. 3 Tampilan Data Tail

4.2 Hasil Pembahasan

4.2.1 Business Understanding



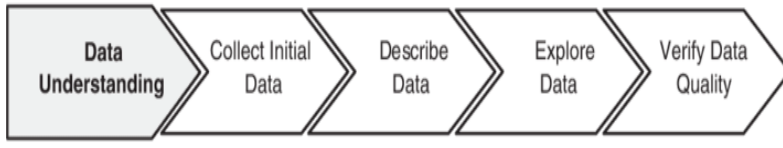
Gambar 4. 4 Langkah Business Understanding

Pada tahapan pertama dilakukan tahapan business understanding dimana pada tahapan business understanding terdapat beberapa Langkah sehingga dapat dilanjutkan ke tahap selanjutnya. Adapun beberapa Langkah tersebut:

- Determine Business Objective* atau menentukan tujuan bisnis. Pada tahap ini peneliti menentukan tujuan apa saja yang akan dilakukan diantaranya yaitu menentukan cara menganalisa karakteristik data, mengetahui perkembangan vaksinasi COVID-19 di DKI Jarta menggunakan algoritma *AdaboostRegressor*, kemudian mengetahui bagaimana kinerja algoritma *AdaboostRegressor* dengan menggunakan RMSE dan MAE.

- b. *Asset situation* atau situasi asset. Pada tahap ini peneliti menentukan faktor lain yang mempertimbangkan penelitian, yaitu berdasarkan website Kementerian Kesehatan Indonesia (kemkes) dikatakan bahwa tercapainya kesuksesan vaksinasi yaitu berdasarkan hasil dosis yang sudah diberikan kepada penduduk, diantaranya berdasarkan data yang peneliti dapatkan yaitu Total vaksinasi yang sudah diberikan yaitu penjumlahan antara vaksin dosis 1 yang sudah diberikan dan vaksin dosis 2 yang sudah diberikan.
- c. *Determine data mining goals* atau menentukan tujuan penambangan data. Pada tahap ini peneliti bertujuan untuk menganalisa ketercapaiannya vaksinasi di Indonesia melalui ketentuan WHO.







4.2.2 Data Understanding



















Gambar 4. 5 Langkah Data Understanding

Tahapan selanjutnya adalah data understanding. Pada tahapan data understanding juga terdapat beberapa langkah sehingga dapat dilanjutkan ke tahap selanjutnya. Adapun beberapa langkah tersebut:

- a. *Collect initial data* atau pengumpulan data awal. Pada tahap ini peneliti mengumpulkan data melalui website satgas COVID-19 DKI Jakarta dan data diolah menggunakan google colaboratory. Data yang dikumpulkan berupa data kumulatif dan dikumpulkan secara harian berdasarkan kelurahan dimana 1 hari terdapat 267 kelurahan dan data yang dikumpulkan bersifat kumulatif. Data yang dikumpulkan dari tanggal 1 Juli sampai 30 November 2021.

FerdyBerlanoP Metrics Done	
 .ipynb_checkpoints	update baru
 Agustus	update baru
 Juli	upload 1
 November	upload 1
 Oktober	upload 1
 September	upload 1

Gambar 4. 6 Kumpulan data perbulan

FerdyBerlanoP upload 1	
..	
 Data Vaksinasi Berbasis Kelurahan (01 Juli 2021).xlsx	upload 1
 Data Vaksinasi Berbasis Kelurahan (02 Juli 2021).xlsx	upload 1
 Data Vaksinasi Berbasis Kelurahan (03 Juli 2021).xlsx	upload 1
 Data Vaksinasi Berbasis Kelurahan (04 Juli 2021).xlsx	upload 1
 Data Vaksinasi Berbasis Kelurahan (05 Juli 2021).xlsx	upload 1
 Data Vaksinasi Berbasis Kelurahan (06 Juli 2021).xlsx	upload 1
 Data Vaksinasi Berbasis Kelurahan (07 Juli 2021).xlsx	upload 1
 Data Vaksinasi Berbasis Kelurahan (08 Juli 2021).xlsx	upload 1
 Data Vaksinasi Berbasis Kelurahan (09 Juli 2021).xlsx	upload 1
 Data Vaksinasi Berbasis Kelurahan (10 Juli 2021).xlsx	upload 1
 Data Vaksinasi Berbasis Kelurahan (11 Juli 2021).xlsx	upload 1
 Data Vaksinasi Berbasis Kelurahan (12 Juli 2021).xlsx	upload 1
 Data Vaksinasi Berbasis Kelurahan (13 Juli 2021).xlsx	upload 1
 Data Vaksinasi Berbasis Kelurahan (14 Juli 2021).xlsx	upload 1
 Data Vaksinasi Berbasis Kelurahan (15 Juli 2021).xlsx	upload 1
 Data Vaksinasi Berbasis Kelurahan (16 Juli 2021).xlsx	upload 1

Gambar 4. 7 Kumpulan Data perhari

- b. *Describe data* atau menjelaskan data. Dari data yang dikumpulkan terdapat beberapa attribute yaitu diantaranya yaitu KODE KELURAHAN, WILAYAH KOTA, KECAMATAN, KELURAHAN, SASARAN, BELUM VAKSIN, JUMLAH DOSIS 1, JUMLAH DOSIS 2, TOTAL VAKSIN DIBERIKAN, LANSIA DOSIS 1, LANSIA DOSIS 2, LANSIA TOTAL VAKSIN DIBERIKAN, PELAYAN PUBLIK DOSIS 1, PELAYAN PUBLIK DOSIS 2, PELAYAN PUBLIK TOTAL VAKSIN DIBERIKAN, GOTONG ROYONG DOSIS 1, GOTONG ROYONG DOSIS 2, GOTONG ROYONG TOTAL VAKSIN DIBERIKAN, TENAGA KESEHATAN DOSIS 1, TENAGA KESEHATAN DOSIS 2, TENAGA KESEHATAN TOTAL VAKSIN DIBERIKAN, TAHAPAN 3 DOSIS 1, TAHAPAN 3 DOSIS 2, TAHAPAN 3 TOTAL VAKSIN DIBERIKAN, REMAJA DOSIS 1, REMAJA DOSIS 2, REMAJA TOTAL VAKSIN DIBERIKAN.

D Data Vaksinasi Berbasis Kelurahan (01 Agustus 2021)									
A	B	C	D	E	F	G	H	I	J
1	DE KELURAHAN	WILAYAH KOTA	KECAMATAN	KELURAHAN	SASARAN	BELUM VAKSIN	JUMLAH DOSIS 1	JUMLAH DOSIS 2	TOTAL VAKSIN DIBERIKAN
2									LANSIA DOSIS 1
3	3172051003	JAKARTA UTARA	PADEMANGAN	ANCOL	8941211	4399496	4541715	1686820	6228535
4	3172051007	JAKARTA BARAT	TAMORA	ANGRE	23947	12155	11792	4268	16060
5	3175041005	JAKARTA TIMUR	KRAMAT JATI	BALE KAMBANG	29031	13727	15454	4905	20559
6	3175031003	JAKARTA TIMUR	JATINGARA	BALI MESTER	29074	18226	10848	3712	14560
7	3175031006	JAKARTA TIMUR	CIPAYUNG	BAMBU APUS	9752	4987	4765	2036	6801
8	3174031002	JAKARTA SELATAN	MAMPANG PRAPATAN	BANGKA	26285	13716	12569	5242	17811
9	3175051002	JAKARTA TIMUR	PASAR REBO	BARU	21566	10232	11334	4089	15423
10	3175041004	JAKARTA TIMUR	KRAMAT JATI	BATU AMPAR	23886	9999	13887	6150	20037
11	3171071002	JAKARTA PUSAT	TANAH ABANG	BENDUNGAN HILIR	47096	28381	19317	7575	26892
12	3175031002	JAKARTA TIMUR	JATINGARA	BIDARA CINA	21494	9895	11399	4662	16261
13	3174101002	JAKARTA SELATAN	PESANGGRAHAN	BINTARO	36689	21668	15021	5556	20577
14	3174011005	JAKARTA SELATAN	TEBET	BUKIT DURI	53098	25851	27247	9951	37198
15	3171041006	JAKARTA PUSAT	SEVEN	BUNJUR	34183	17980	16203	6235	22438
16	3175061007	JAKARTA TIMUR	CAKUNG	CAKUNG BARAT	18144	8318	9626	3815	13641
17	3175061004	JAKARTA TIMUR	CAKUNG	CAKUNG TIMUR	57724	25769	31955	7383	39338
18	3175061007	JAKARTA TIMUR	KRAMAT JATI	KAWANG	58287	27205	31082	10319	41401
19	317501008	JAKARTA TIMUR	CIPAYUNG	CEGER	32762	17888	14874	5940	20814
20	3171031006	JAKARTA PUSAT	KEMAYORAN	CEMPAKA BARU	18932	9316	9616	3489	13105
					32581	17199	15382	5622	21004

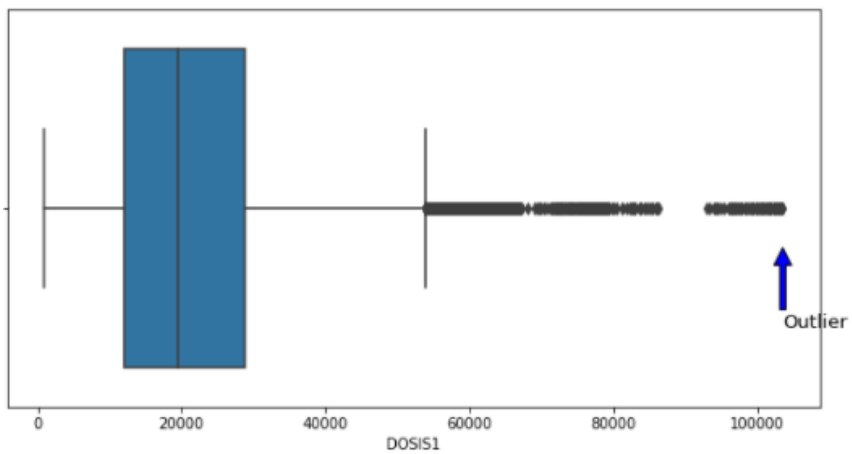
Gambar 4. 8 Data dalam bentuk Excel

- c. *Explore data* atau menjelajahi data. Melihat data attribute mana saja yang akan digunakan dan melihat data jika memiliki kekosongan pada hari atau attribute atau nilai yang kosong ataupun data outlier.

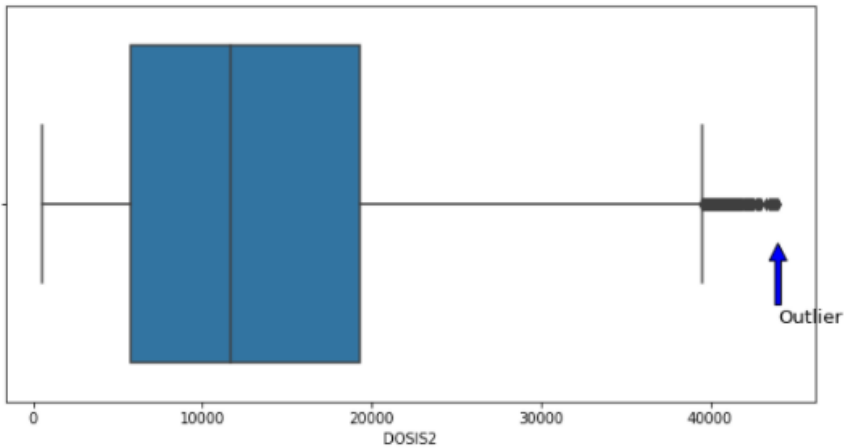
```
df.isna().sum()
```

```
WILAYAH KOTA      0
KECAMATAN         0
KELURAHAN         0
SASARAN           0
BELUM VAKSIN      0
TOTAL VAKSIN\ndIBERIKAN  0
dtype: int64
```

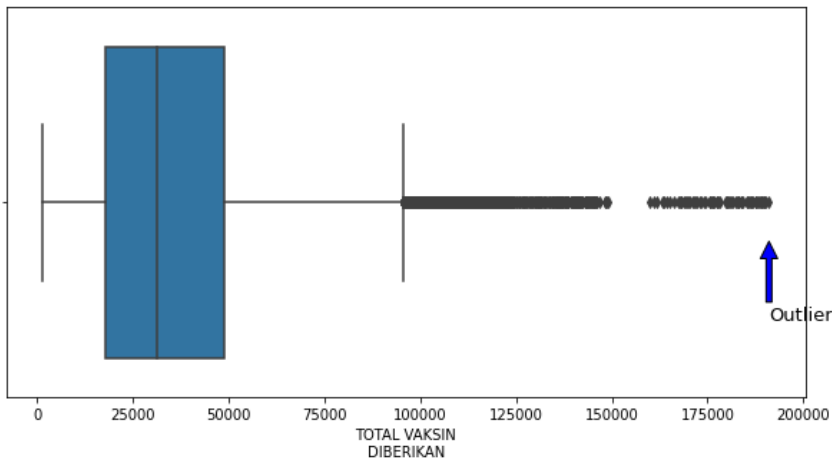
Gambar 4. 9 Melihat nilai kosong pada data



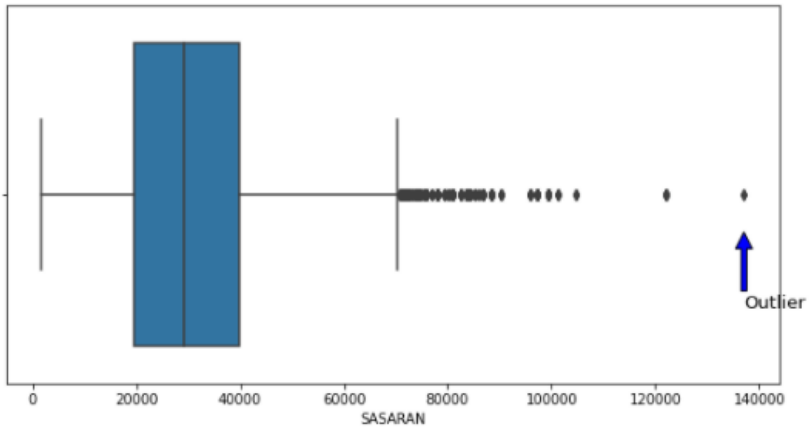
Gambar 4. 10 Data Outlier Tabel Dosis 1



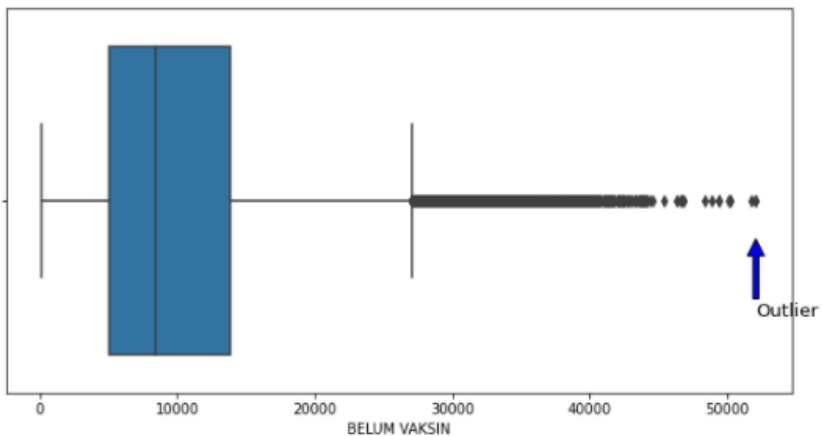
Gambar 4. 11 Data Outlier Tabel Dosis 2



Gambar 4. 12 Data Outlier Tabel Total Vaksin Diberikan



Gambar 4. 13 Data Outlier Tabel Sasaran



Gambar 4. 14 Data Outlier Tabel Belum Vaksin

- d. *Verify data quality* atau memverifikasi kualitas data. Data yang di dapatkan tidak lengkap. Beberapa di setiap bulannya ada beberapa data yang kosong atau tidak data di tanggal tertentu. Contohnya di bulan agustus tidak ada data tanggal 28 Agustus 2021.

4.2.3 Data Preparation



Gambar 4. 15 Langkah Data Preparation

Tahapan selanjutnya adalah data preparation. Pada tahapan preparation juga terdapat berbagai langkah yang harus dilakukan supaya tidak terjadi permasalahan pada langkah selanjutnya. Adapun beberapa langkah tersebut:

- a. Sesaat sebelum memilih data peneliti menggabungkan file yang terdapat pada gambar (gambar github) yaitu menggabungkannya menjadi satu file.
- b. Select data atau memilih/menyeleksi data. Pada tahap ini dari beberapa attribute hanya digunakan 6 atribut diantaranya yaitu KOTA, KECAMATAN, KELURAHAN, SASARAN, BELUM VAKSIN, DOSIS 1, DOSIS 2, TOTAL. Pemilihan attribute ini didasarkan melalui website kemkes pada tanggal 14 November 2021 mengatakan bahwa vaksinasi di Indonesia berhasil mencapai target WHO mencapai 215,6 juta dosis vaksin diberikan cakupannya yaitu vaksin dosis 1, vaksin dosis 2, vaksin dosis 3 atau booster. Namun dikarenakan atribut vaksin dosis 3 tidak tersedia, jadi peneliti menggunakan vaksin dosis 1 dan dosis 2 dan total vaksin yang sudah diberikan yaitu penjumlahan antara vaksin dosis 1 dan dosis 2.

TANGGAL	KOTA	KECAMATAN	KELURAHAN	SASARAN	BELUM VAKSIN	DOSIS1	DOSIS2	TOTAL
2021-07-01	JAKARTA UTARA	PADEMANGAN	ANCOL	20393	13272	7114	3287	10401
2021-07-01	JAKARTA BARAT	TAMBORA	ANGKE	25785	16477	9299	3221	12520
2021-07-01	JAKARTA TIMUR	KRAMAT JATI	BALE KAMBANG	25158	18849	6301	2644	8945
2021-07-01	JAKARTA TIMUR	JATINEGARA	BALI MESTER	8683	5743	2937	1517	4454
2021-07-01	JAKARTA TIMUR	CIPAYUNG	BAMBU APUS	22768	15407	7357	3985	11342
...
2021-11-30	JAKARTA TIMUR	MATRAMAN	UTAN KAYU SELATAN	33511	5753	27758	23688	51446
2021-11-30	JAKARTA TIMUR	MATRAMAN	UTAN KAYU UTARA	28169	5448	22721	19074	41795
2021-11-30	JAKARTA PUSAT	KEMAYORAN	UTAN PANJANG	29208	5515	23693	19731	43424
2021-11-30	JAKARTA UTARA	TANJUNG PRIOK	WARAKAS	44401	7828	36573	30403	66976
2021-11-30	JAKARTA BARAT	GROGOL PETAMBURAN	WIJAYA KUSUMA	38934	7741	31193	27022	58215

37380 rows x 8 columns

Gambar 4. 16 Seleksi Data

- c. Clean data atau membersihkan data. Saat membersihkan data peneliti menggunakan beberapa metode diantaranya yaitu menggunakan interpolasi dan menggunakan outlier. Adapun intepolasi yang digunakan adalah intepolasi linier. Selain menggunakan interpolasi linier peneliti juga menggunakan metode outlier. Outlier sendiri yaitu objek yang menyimpang dari objek lainnya. Metode outlier yang digunakan pada penelitian ini adalah metode inter-quartile range (IQR). Untuk menentukan metode IQR sendiri yaitu dengan cara menentukan batas atas dan batas bawah terlebih dahulu. Untuk menentukan batas bawah yaitu buatlah variable dengan nama q1 terlebih dahulu lalu kita mencari tahu percentil dari 25% attribute yang digunakan, sedangkan untuk batas atas buatlah variable dengan nama q3 lalu mencari tahu dari percentil 75% attribute yang digunakan. Jika sudah di dapatkan maka hitung IQR dengan cara mengurangi nilai q3 dengan q1. Setelah didapatkan kita melanjutkan untuk membuat nilai batas bawah dengan cara q1 dikurangi 1,5 dikalikan dengan nilai IQR. dan batas atas dengan cara q3 ditambah 1,5 dikali IQR. maka didapatkanlah nilai batas atas dan batas bawahnya.

```
df = df.interpolate(method='linear')
df
```

TANGGAL	KOTA	KECAMATAN	KELURAHAN	SASARAN	BELUM VAKSIN	DOSIS1	DOSIS2	TOTAL
2021-07-01	JAKARTA UTARA	PADEMANGAN	ANCOL	20393	13272	7114	3287	10401
2021-07-01	JAKARTA BARAT	TAMBORA	ANGKE	25785	16477	9299	3221	12520
2021-07-01	JAKARTA TIMUR	KRAMAT JATI	BALE KAMBANG	25158	18849	6301	2644	8945
2021-07-01	JAKARTA TIMUR	JATINEGARA	BALI MESTER	8683	5743	2937	1517	4454
2021-07-01	JAKARTA TIMUR	CIPAYUNG	BAMBU APUS	22768	15407	7357	3985	11342
...
2021-11-30	JAKARTA TIMUR	MATRAMAN	UTAN KAYU SELATAN	33511	5753	27758	23688	51446
2021-11-30	JAKARTA TIMUR	MATRAMAN	UTAN KAYU UTARA	28169	5448	22721	19074	41795
2021-11-30	JAKARTA PUSAT	KEMAYORAN	UTAN PANJANG	29208	5515	23693	19731	43424
2021-11-30	JAKARTA UTARA	TANJUNG PRIOK	WARAKAS	44401	7828	36573	30403	66976
2021-11-30	JAKARTA BARAT	GROGOL PETAMBURAN	WIJAYA KUSUMA	38934	7741	31193	27022	58215

Gambar 4. 17 Interpolasi

```
q1 = dt['DOSIS1'].quantile(0.25)
q3 = dt['DOSIS1'].quantile(0.75)
iqr=q3-q1
lower_limit= q1 - 1.5 * iqr
upper_limit= q3 + 1.5 * iqr
dt=dt[(dt['DOSIS1']<lower_limit)|(dt['DOSIS1']>upper_limit)]
dt
```

Gambar 4. 18 Pembersihan Outlier Tabel Dosis1 Menggunakan Metode IQR

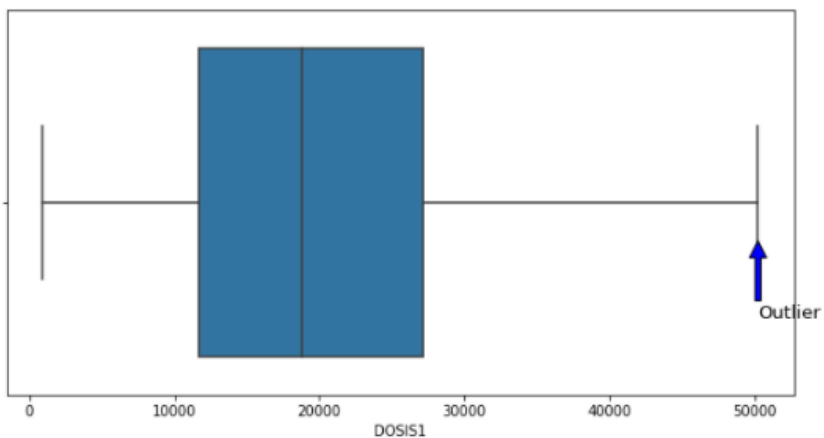
```
q1 = dt['DOSIS2'].quantile(0.25)
q3 = dt['DOSIS2'].quantile(0.75)
iqr=q3-q1
lower_limit= q1 - 1.5 * iqr
upper_limit= q3 + 1.5 * iqr
dt=dt[(dt['DOSIS2']<lower_limit)|(dt['DOSIS2']>upper_limit)]
dt
```

Gambar 4. 19 Pembersihan Outlier Tabel Dosis2 Menggunakan Metode IQR

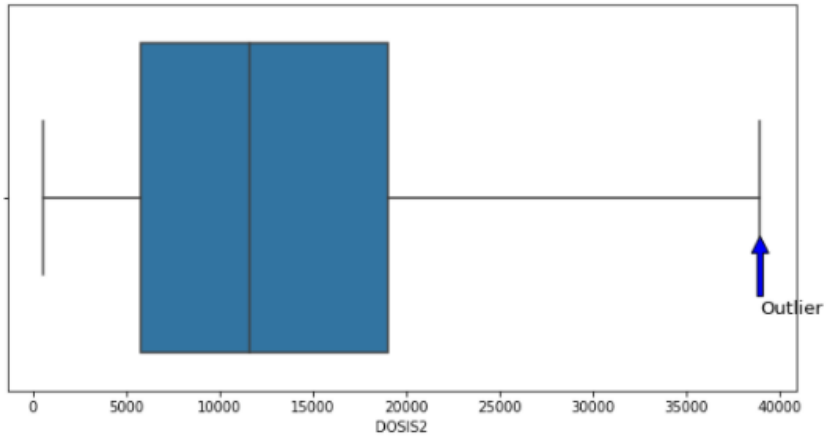

```
q1 = dt['TOTAL'].quantile(0.25)
q3 = dt['TOTAL'].quantile(0.75)
iqr=q3-q1
lower_limit= q1 - 1.5 * iqr
upper_limit= q3 + 1.5 * iqr
dt=dt[(dt['TOTAL']<lower_limit)|(dt['TOTAL']<upper_limit)]
dt
```

Gambar 4. 20 Pembersihan Outlier Tabel Total Vaksin diberikan Menggunakan Metode IQR

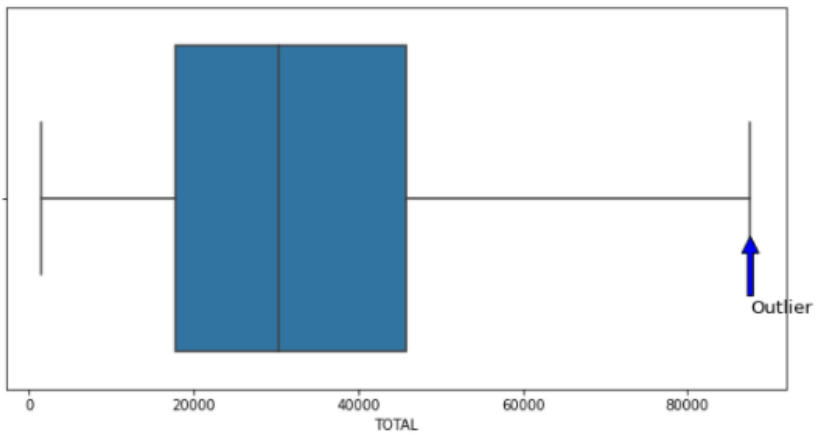
Membersihkan data outlier dilakukan hingga tidak ada outlier yang tersisa.



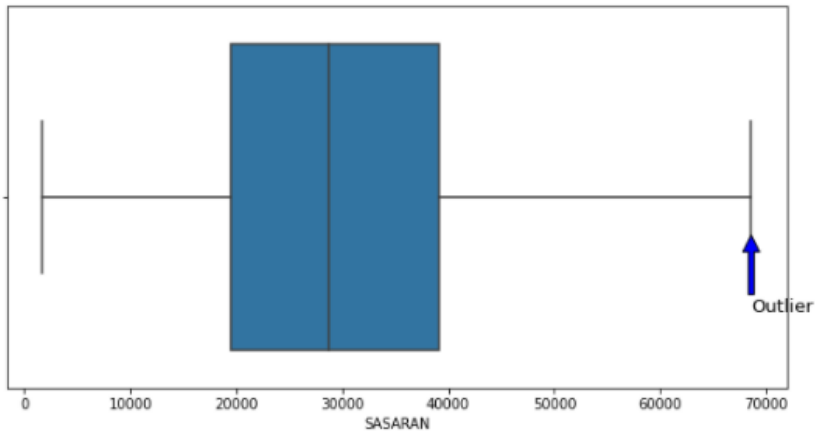
Gambar 4. 21 Data Oulier Setelah dibersihkan Pada Tabel Dosis1



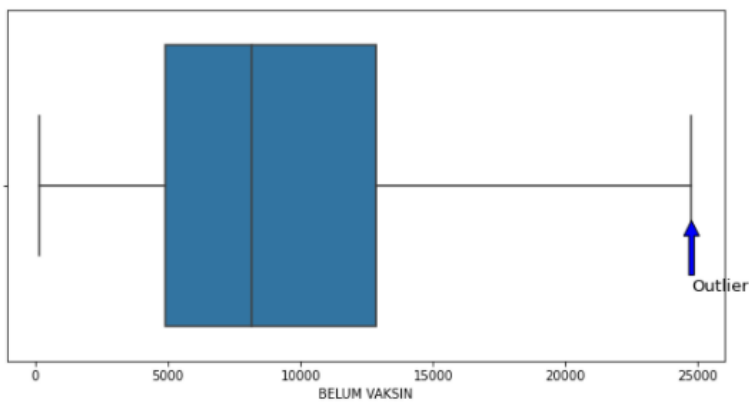
Gambar 4. 22 Data Oulier Setelah dibersihkan Pada Tabel Dosis2



Gambar 4. 23 Data Oulier Setelah dibersihkan Pada Tabel Total Vaksin diberikan



Gambar 4. 24 Data Oulier Setelah dibersihkan Pada Tabel Sasaran



Gambar 4. 25 Data Oulier Setelah dibersihkan Pada Tabel Belum Vaksin

Setelah data dibersihkan dengan metode outliers data yang tersisa berjumlah 32747

- d. Format data. Pada proses ini peneliti melakukan format data menggunakan fitur pada library yang disediakan oleh scikit-learn yaitu “Label Encoder”. Fungsinya sendiri adalah mengubah string menjadi numerik. Dimana nama dari “KOTA,

KECAMATAN, KELURAHAN” diubah menjadi numerik untuk mempermudah proses pemodelan.

TANGGAL	KOTA	KECAMATAN	KELURAHAN	SASARAN	BELUM VAKSIN	DOSIS2	TOTAL
2021-07-01	4	28	0	289	10803	1942	3220
2021-07-01	0	40	1	440	12887	1901	4088
2021-07-01	3	23	2	420	14007	1450	2619
2021-07-01	3	11	3	96	4529	588	1083
2021-07-01	3	5	4	354	12271	2517	3588
...
2021-11-30	3	26	246	628	4539	15508	20228
2021-11-30	3	26	247	499	4249	13112	17295
2021-11-30	1	18	248	529	4310	13469	17845
2021-11-30	4	42	249	805	6427	18009	23751
2021-11-30	0	9	250	745	6347	17056	22207

Gambar 4. 26 Mengubah Data String Menjadi Data Numerik

4.2.4 Modelling



Gambar 4. 27 Langkah Modelling

Tahapan selanjutnya adalah Modelling. Pada tahapan modelling juga terdapat berbagai langkah yang harus dilakukan supaya tidak terjadi permasalahan pada langkah selanjutnya. Adapun beberapa langkah tersebut:

- Select modelling technique atau memilih teknik pemodelan. Setelah selesai menyiapkan data peneliti menentukan pemodelan yaitu regressor dengan dengan teknik AdaBoostRegressor.
- Generate test design atau menghasilkan desain pengujian. Kemudian dilakukan pemisahan dataset menjadi 2 yaitu train dan

test set dengan perbandingan 70:30. 70% untuk train set dan 30% untuk test set. Dengan random state adalah membagi data secara acak. Jadi jika proses running diulangi maka akan membuat hasil yg didapatkan berubah-ubah.

```
X_train,X_test,y_train,y_test=train_test_split(X,y,random_state=100,test_size=.3)
```

Gambar 4. 28 Split Data

- c. Build model atau membangun model. Setelah selesai membagi data menjadi 2 kelompok selanjutnya lakukan dengan pemodelan.

```
model=AdaBoostRegressor()  
model.fit(X_train,y_train)
```

```
AdaBoostRegressor()
```

Gambar 4. 29 Penggunaan Model

- d. Assess model atau menilai model. Setelah selesai menerapkan model maka cari nilai yang didapatkan setelah menerapkan model tersebut. Nilai yang didapatkan adalah 98% dengan nilai korelasi 99%

```
model.score(X_test,y_test)
```

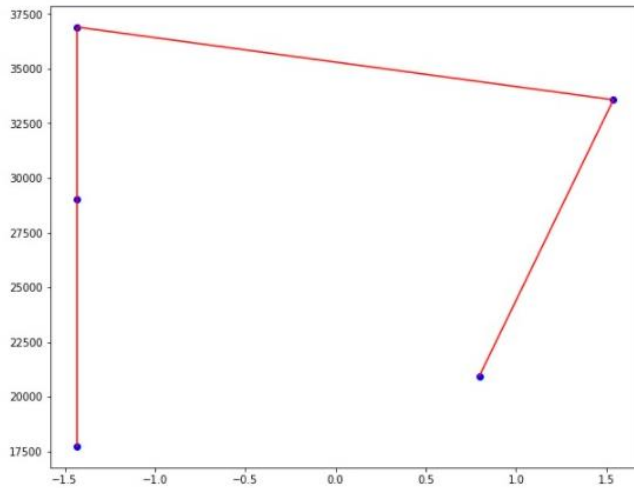
```
0.9861823360178884
```

```
print('Correlation: ', math.sqrt(model.score(X_test,y_test)))
```

```
Correlation: 0.9930671357052797
```

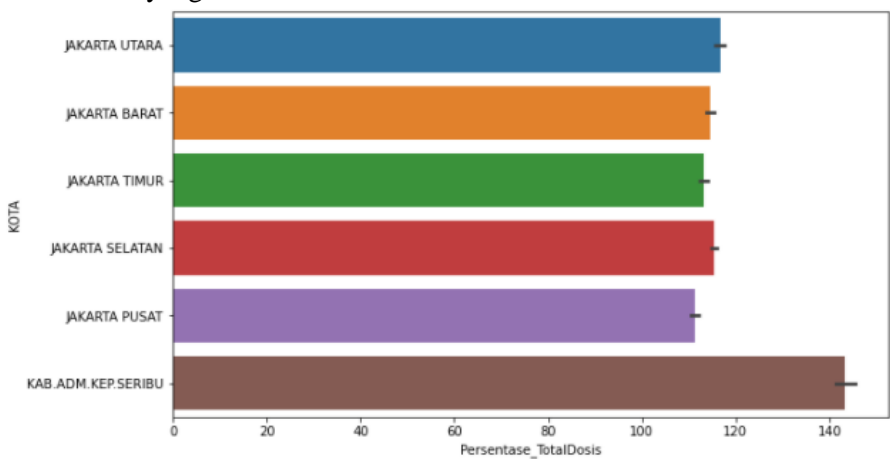
Gambar 4. 30 Nilai Model dan Korelasi

- e. Visualisasi algoritma adaboost regressor. Setelah selesai melakukan pemodelan maka kita buat visualisasinya Adapun hasilnya sebagai berikut.

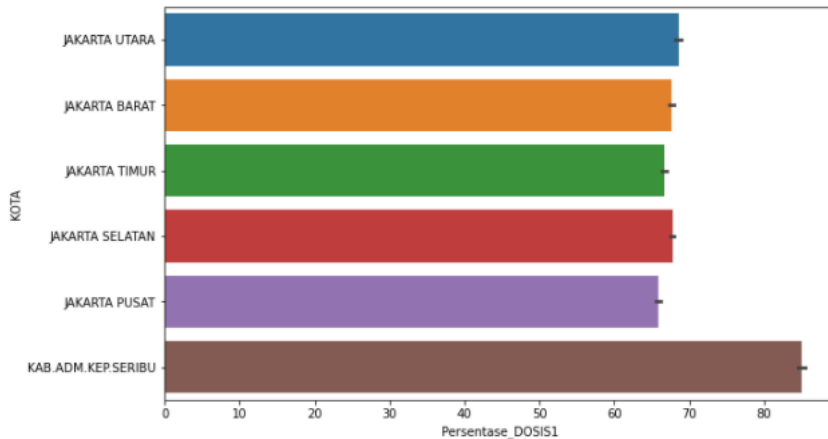


Gambar 4. 31 Visualisais Algoritma AdaBoost Regressor

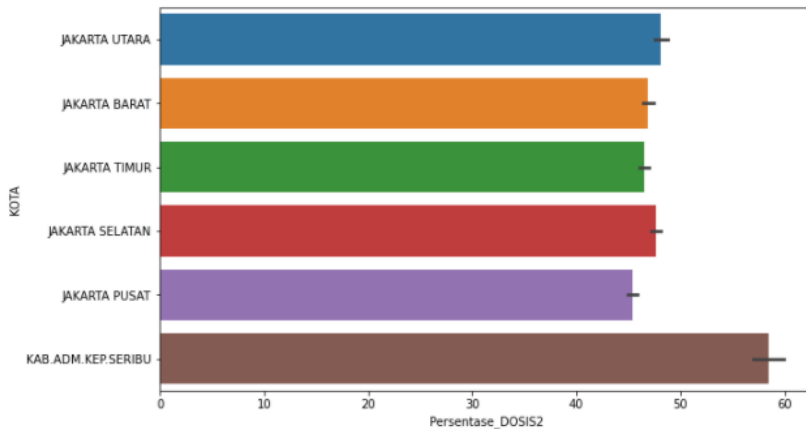
- f. Hasil perkembangan vaksinasi
Berdasarkan info Kementrian Kesehatan Indonesia dikatakan tercapai vaksinasi adalah jika sudah mencapai lebih kurang 80% dosis yang sudah disebar baik dosis1 dan dosis2.



Gambar 4. 32 Perkembangan vaksinasi Total Dosis yang Sudah Diberikan



Gambar 4. 33 Perkembangan Vaksinasi Dosis1 yang Sudah Diberikan



Gambar 4. 34 Perkembangan Vaksinasi Dosis2 yang Sudah Diberikan

4.2.5 Evaluation

Tahapan selanjutnya adalah evaluation. Pada tahapan evalution juga terdapat berbagai langkah yang harus dilakukan supaya tidak terjadi permasalahan pada langkah selanjutnya. Adapun beberapa langkah tersebut:

- a. Evaluate results atau evaluasi hasil. Setelah selesai maka dilakukan peninjauan ulang terhadap model dengan

menggunakan MAE dan RMSE dengan nilai MAE 1040.36 dan RMSE 1240.83

```
print('Mean Absolute Error (MAE):', metrics.mean_absolute_error(y_test,y_prediksi))  
print('Root Mean Squared Error (RMSE):', np.sqrt(metrics.mean_squared_error(y_test,y_prediksi)))
```

```
Mean Absolute Error (MAE): 1040.3676511548688  
Root Mean Squared Error (RMSE): 1240.8315038661776
```

Gambar 4. 35 Nilai MAE dan RMSE

Bab V

Kesimpulan dan Saran

5.1 Kesimpulan

Untuk mencegah penyakit yang sangat menular dari COVID-19, proses vaksinasi yang aman dan efektif adalah wajib. Proses vaksinasi dapat menghilangkan rasa takut terinfeksi saat terpapar dengan orang lain. Pada penelitian Analisis Ketercapaian Vaksinasi Terhadap Penyebaran COVID-19 di DKI Jakarta menggunakan model AdaBoost Regressor. Sebelum menerapkan model data dibersihkan dengan metode outlier yang di cek menggunakan *boxplot* lalu data di uji menggunakan model Adaboost dengan nilai akurasi senilai 98% dan nilai korelasi 99%. Namun sangat disayangkan nilai dari MAE dan RMSE masih jauh dari nilai nol yaitu MAE dengan nilai yang didapatkan yaitu 1040.3676 dan nilai RMSE yaitu 1240.8315.

5.2 Saran

Pada sistem ini tentunya ada beberapa kekurangan dan disarankan menggunakan metode yang mungkin lebih baik dan pada sistem ini tentunya ada beberapa saran yang dapat dilakukan untuk pengembangan selanjutnya. Memperluas ruang lingkup yang dapat diterapkan salah satunya adalah melihat perkembangan ekonomi dari dampak vaksinasi atau penurunan tingkat kasus positif COVID-19 dengan dengan model yang lebih baik dan tingkat akurasi yang sama atau lebih baik.

Daftar Pustaka

- [1] C. Huang *et al.*, “Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China,” *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020, doi: 10.1016/S0140-6736(20)30183-5.
- [2] P. Wu *et al.*, “Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020,” *Eurosurveillance*, vol. 25, no. 3, pp. 1–6, 2020, doi: 10.2807/1560-7917.ES.2020.25.3.2000044.
- [3] P. K. VIPIN M VASHISHTHA, “Emergency use authorisation of Covid-19 vaccines: An ethical conundrum,” *Indian J. Med. Ethics* 2, vol. VI, no. 1, pp. 20–22, 2020, doi: 10.1016/j.
- [4] W. H. Organization, “COVID-19 advice for the public: Getting vaccinated,” *World Health Organization*, 2020. .
- [5] A. Dogan and D. Birant, “Machine learning and data mining in manufacturing,” *Expert Syst. Appl.*, vol. 166, p. 114060, 2021, doi: 10.1016/j.eswa.2020.114060.
- [6] V. C. Parro, M. L. M. Lafetá, F. Paít, F. B. Ipólito, and T. N. Toporcov, “Predicting COVID-19 in very large countries: The case of Brazil,” *PLoS One*, vol. 16, no. 7 July, pp. 1–15, 2021, doi: 10.1371/journal.pone.0253146.
- [7] M. A. Quiroz-Juárez, A. Torres-Gómez, I. Hoyo-Ulloa, R. D. J. de León-Montiel, and A. B. U’Ren, “Identification of high-risk COVID-19 patients using machine learning,” *PLoS One*, vol. 16, no. 9 September, pp. 1–21, 2021, doi: 10.1371/journal.pone.0257234.
- [8] Y. Hao, T. Xu, H. Hu, P. Wang, and Y. Bai, “Prediction and analysis of Corona virus disease 2019,” *PLoS One*, vol. 15, no. 10 October, 2020, doi: 10.1371/journal.pone.0239960.
- [9] M. E. Elkin and X. Zhu, “Understanding and predicting COVID-19 clinical trial completion vs. cessation,” *PLoS One*, vol. 16, no. 7 July, pp. 1–25, 2021, doi: 10.1371/journal.pone.0253789.
- [10] L. Tseng *et al.*, “Predicting Poor Outcome of COVID-19 Patients on the Day of Admission with the COVID-19 Score,” *Crit. Care Res.*

Pract., vol. 2021, 2021, doi: 10.1155/2021/5585291.

- [11] P. Mars, J. R. Chen, and R. Nambiar, "Regression Analysis of COVID-19 using Machine Learning Algorithms," *Learn. Algorithms*, no. Icosec, pp. 65–71, 2018, doi: 10.1201/9781351073974.
- [12] Y. Zoabi, S. Deri-Rozov, and N. Shomron, "Machine learning-based prediction of COVID-19 diagnosis based on symptoms," *npj Digit. Med.*, vol. 4, no. 1, pp. 1–5, 2021, doi: 10.1038/s41746-020-00372-6.
- [13] J. V. Lazarus *et al.*, "A global survey of potential acceptance of a COVID-19 vaccine," *Nat. Med.*, vol. 27, no. 2, pp. 225–228, 2021, doi: 10.1038/s41591-020-1124-9.
- [14] T. M. Cook and J. V. Roberts, "Impact of vaccination by priority group on UK deaths, hospital admissions and intensive care admissions from COVID-19," *Anaesthesia*, vol. 76, no. 5, pp. 608–616, 2021, doi: 10.1111/anae.15442.
- [15] S. Dutta, U. Mukherjee, and S. Kumar Bandyopadhyay, "Pharmacy Impact on Vaccination Progress Using Machine Learning Approach," 2021, doi: 10.20944/preprints202106.0533.v1.
- [16] A. Fridman, R. Gershon, and A. Gneezy, "COVID-19 and vaccine hesitancy: A longitudinal study," *PLoS One*, vol. 16, no. 4 April, pp. 1–12, 2021, doi: 10.1371/journal.pone.0250123.
- [17] P. L. Espinheira, L. C. M. Silva, and F. Cribari-Neto, "Bias and variance residuals for machine learning nonlinear simplex regressions," *Expert Syst. Appl.*, vol. 185, no. May 2020, p. 115656, 2021, doi: 10.1016/j.eswa.2021.115656.
- [18] C. Srisa-An, "Guideline of Collinearity - Avoidable Regression Models on Time-series Analysis," in *2021 2nd International Conference on Big Data Analytics and Practices (IBDAP)*, 2021, pp. 28–32, doi: 10.1109/IBDAP52511.2021.9552165.
- [19] F. Xiao, Y. Wang, L. He, H. Wang, W. Li, and Z. Liu, "Motion Estimation From Surface Electromyogram Using Adaboost Regression and Average Feature Values," *IEEE Access*, vol. 7, pp. 13121–13134, 2019, doi: 10.1109/ACCESS.2019.2892780.
- [20] H. Jiang, W. Zheng, L. Luo, and Y. Dong, "A two-stage minimax concave penalty based method in pruned AdaBoost ensemble," *Appl. Soft Comput. J.*, vol. 83, p. 105674, 2019, doi:

10.1016/j.asoc.2019.105674.

- [21] X. Zhu, P. Zhang, and M. Xie, "A Joint Long Short-Term Memory and AdaBoost regression approach with application to remaining useful life estimation," *Meas. J. Int. Meas. Confed.*, vol. 170, no. September, p. 108707, 2021, doi: 10.1016/j.measurement.2020.108707.
- [22] D. S. K. Karunasingha, "Root mean square error or mean absolute error? Use their ratio as well," *Inf. Sci. (Ny)*, vol. 585, pp. 609–629, Mar. 2022, doi: 10.1016/J.INS.2021.11.036.
- [23] J. Hao and T. K. Ho, "Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language," *J. Educ. Behav. Stat.*, vol. 44, no. 3, pp. 348–361, 2019, doi: 10.3102/1076998619832248.
- [24] "Matplotlib." <https://matplotlib.org/>.
- [25] "Seaborn." <https://seaborn.pydata.org/>.
- [26] "Pandas." <https://pandas.pydata.org/>.
- [27] "Numpy." <https://numpy.org/>.
- [28] "Scikit-learn." [Online]. Available: <https://scikit-learn.org/>.
- [29] N. N. R. R. Suri, N. Murty, and M. G. Athithan, *Outlier Detection: Techniques and Applications - A Data Mining Perspective*. 2019.
- [30] G. Huang, "Missing data filling method based on linear interpolation and lightgbm," *J. Phys. Conf. Ser.*, vol. 1754, no. 1, 2021, doi: 10.1088/1742-6596/1754/1/012187.
- [31] D. Arimbi, A. Afifuddin, and R. W. Sekarsari, "Dampak Pengembangan Dan Pembangunan Alun-Alun Kota Wisata Batu terhadap Jumlah Sampah di Kawasan Alun-Alun Kota Batu (Studi Kasus: Manajemen Alun ...," *Respon Publik*, vol. 14, no. 2, pp. 111–118, 2020, [Online]. Available: <http://www.riset.unisma.ac.id/index.php/rpp/article/view/8078>.
- [32] P. C. Ncr *et al.*, "Crisp-Dm," *SPSS inc*, vol. 78, pp. 1–78, 2000, [Online]. Available: <http://www.crisp-dm.org/CRISPWP-0800.pdf>.