

ANALISIS JUMLAH KEMATIAN AKIBAT COVID-19 BERDASARKAN PROVINSI DENGAN METODE EXPLORATORY DATA ANALYSIS (EDA)

Analysis The Number Of Deaths Due To Covid-19 By Province With Exploratory Data Analysis (EDA) Method

Akmal Ihab Syauqi¹, Muhammad Faiq Ammar², Haikal Nuril Abiyit³, Rafi Jauhari⁴, Muhammad Afif Rohman Muzaky⁵

Prodi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Jember¹²³⁴⁵

222410101034@mail.unej.ac.id¹, 222410101039@mail.unej.ac.id²,
222410101058@mail.unej.ac.id³, 222410101087@mail.unej.ac.id⁴,
222410101095@mail.unej.ac.id⁵

ABSTRAK: Pada awal tahun 2020 hingga akhir 2022, jumlah kematian yang disebabkan oleh wabah penyakit Covid-19 sangatlah banyak. Banyak upaya yang dilakukan pemerintah untuk mengendalikan laju penyebaran Covid-19. Tujuan dari penelitian yang kami lakukan adalah untuk menjelaskan dan memudahkan dalam menganalisis data jumlah kematian yang disebabkan oleh Covid-19 berdasarkan provinsi. Disini kami menggunakan metode *Exploratory Data Analysis* (EDA) untuk melakukan visualisasi data dan Data Cleaning untuk membersihkan kolom atau data outliers yang tidak dibutuhkan dalam penelitian data lanjutan mengenai jumlah kematian akibat Covid-19. Dengan menggunakan metode ini Diharapkan dengan adanya penelitian ini, dapat diketahui provinsi mana saja dengan dampak kematian terbanyak yang disebabkan oleh Covid-19.

Kata kunci: COVID-19, Penyakit, Metode Exploratory Data Analysis (EDA), Visualisasi Data

ABSTRACT: In early 2020 to late 2022, the number of deaths caused by the Covid-19 disease outbreak is undeniably large. Many efforts have been made by the government to control the spread of Covid-19. Our research aims to explain and facilitate data on the number of deaths caused by Covid-19 by province. Here we use the *Exploratory Data Analysis* (EDA) to perform data visualization and Data Cleaning methods to clean columns or outliers that are unnecessary in further data research on the number of deaths due to Covid-19. Using this method, it is expected that with this research, we can find out which provinces have the most deaths caused by Covid-19.

Keywords: COVID-19, Disease, Exploratory Data Analysis (EDA) Method, Data Visualization

1. PENDAHULUAN

Pada tahun 2020, digemparkan dengan kehadiran virus baru yaitu coronavirus jenis baru yang bernama SARS-CoV-2, yang kemudian diberi nama Coronavirus Disease 2019 (COVID-19). COVID-19 adalah penyakit yang menyebabkan gangguan pernapasan dan radang paru. Gejala klinis yang muncul beragam, mulai dari seperti gejala flu.

Pemerintah Indonesia terus membuat kebijakan untuk mengendalikan laju pertumbuhan Covid-19, seperti menggunakan alat pendeteksi, pengukur suhu untuk memeriksa suhu tubuh, bahkan mengerahkan banyak tenaga kerja untuk membantu pasien yang terinfeksi Covid-19.

Dilansir dari databokskatadata.com sampai tanggal 14 Agustus 2023, Indonesia menduduki peringkat dua dengan total kematian terbanyak yang diakibatkan oleh Covid-19.

Menurut infesiemergingkemkes.com sebanyak 161.918 orang yang meninggal akibat Covid-19. Data tersebut dicatat oleh pemerintah provinsi.

Menurut Laporcovid19.org terdapat selisih data angka kematian Covid-19 berdasarkan data Covid-19 per provinsi dengan versi publikasi pemerintah pusat. Terhitung sebesar 10.345 Jiwa perbedaan. Hal tersebut dikarenakan banyaknya data null dalam data pemerintah.

Tujuan dari penelitian yang kami lakukan adalah untuk menjelaskan dan memudahkan dalam menganalisis data jumlah kematian yang disebabkan oleh Covid-19 berdasarkan provinsi. Disini kami menggunakan metode *Exploratory Data Analysis* (EDA) untuk melakukan visualisasi data dan Data Cleaning untuk membersihkan kolom atau data outliers yang tidak dibutuhkan dalam penelitian data lanjutan mengenai jumlah kematian akibat Covid-19. Dengan menggunakan metode ini diharapkan dengan adanya penelitian ini, dapat diketahui provinsi mana saja dengan dampak kematian terbanyak yang disebabkan oleh Covid-19.

Hasil dari penelitian ini adalah masih terdapat banyak data null dalam data kematian akibat Covid-19 yang diupload versi pemerintah pusat. Setelah menggunakan metode Exploratory Data Analysis (EDA) dan Visualisasi Data (VD), semua data data null hilang.

2. METODE PENELITIAN

Penelitian ini menggunakan metode Exploratory Data Analysis (EDA) dan Visualisasi Data (VD).



Gambar 1. Diagram Alur Penelitian

2.1 Studi Literatur

Dalam studi literatur ini kami mengumpulkan data pustaka dengan membaca, mencatat, serta mengolah bahan penelitian

2.2 Data Acquisition

Data acquisition adalah proses pengumpulan informasi untuk menganalisa suatu gejala fisik menggunakan komputer. Data yang kita akuisisi pada artikel ini adalah Jumlah kematian yang disebabkan oleh COVID-19 di Indonesia

2.3 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) adalah proses menganalisis dan menampilkan data bertujuan mendapatkan pemahaman yang lebih baik tentang wawasan dari data. Langkah-langkah dalam melakukan EDA antara lain :

1. Mencari wawasan terhadap kumpulan data
2. Mengungkap struktur data

3. Mengekstrak variabel yang penting
4. Mendeteksi Outliers
5. Melakukan uji asumsi
6. Mengembangkan model
7. Menentukan faktor yang optimal.

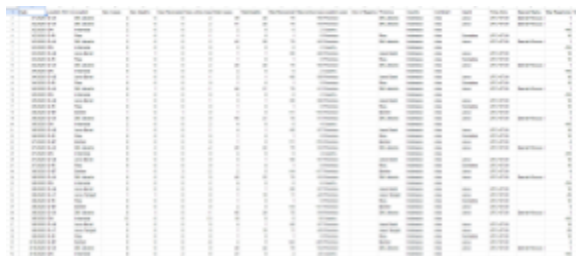
Peran utama EDA adalah untuk mengeksplorasi data secara terbuka, dan grafik bertujuan memperkuat analisis yang dilakukan.

2.4 Visualisasi Data

Visualisasi data adalah proses penggunaan elemen visual seperti diagram, grafik, atau peta untuk merepresentasikan dan menampilkan data. Visualisasi data menerjemahkan data yang kompleks, ber volume tinggi menjadi representasi visual yang lebih mudah dipahami.

3. HASIL DAN PEMBAHASAN

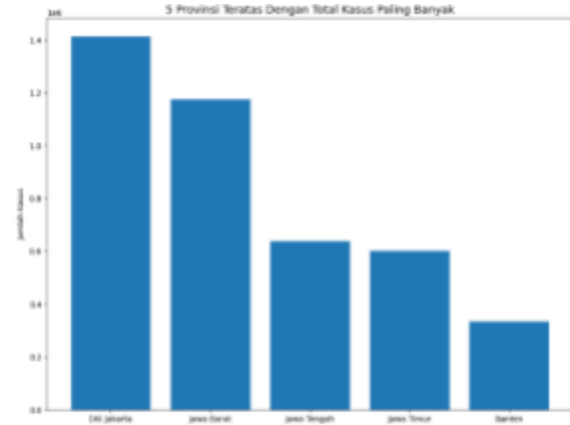
Berikut adalah dataset yang kami gunakan pada penelitian kami :



Provinsi	Kasus Baru	Kumulatif Kasus	Kematian	Recovery
DKI Jakarta	14000	140000	1500	125000
Jawa Barat	11000	110000	1000	99000
Jawa Tengah	6000	60000	600	54000
Jawa Timur	6000	60000	600	54000
Banten	3000	30000	300	27000

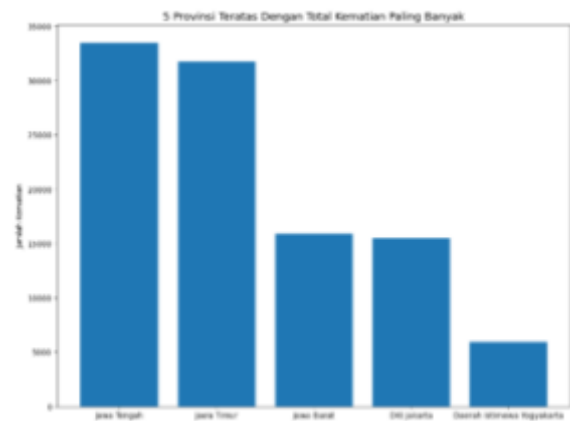
Gambar 1. Dataset COVID-19 di Indonesia

Hasil yang kita dapatkan adalah grafik batang yang berisikan lima provinsi yang memiliki jumlah kasus terbanyak dan jumlah kematian terbanyak. Gambar 2 memaparkan lima provinsi dengan jumlah kasus Covid-19 terbanyak. DKI Jakarta (1.4 juta), Jawa Barat (1.1 Juta), Jawa Tengah (636 ribu), Jawa Timur (601 ribu), dan Banten (333 ribu). Provinsi dengan jumlah kasus Covid-19 terbanyak adalah DKI Jakarta.



Gambar 2. Provinsi dengan kasus COVID terbanyak

Gambar 3 memaparkan lima provinsi dengan jumlah kematian akibat Covid-19 terbanyak. Jawa Tengah (33.4 ribu), Jawa Timur (31.7 ribu), Jawa Barat (15.9 ribu), DKI Jakarta (15.5 ribu), dan Daerah Istimewa Yogyakarta (5.9 ribu). Provinsi dengan jumlah kematian akibat Covid-19 terbanyak adalah Jawa Tengah.



Gambar 3. Provinsi dengan jumlah total kematian paling banyak

Dari tiga gambar diatas kita dapat melihat visualisasi dari data Kasus Covid-19 berdasarkan provinsi dan dapat kita mengambil kesimpulan dengan melakukan analisis jumlah kematian akibat covid-19 berdasarkan provinsi.

-----LinearRegression()-----

MAE: 4.742804731751494

MSE: 337.06177009388784

RMSE: 18.35924208930989

R2 Square 0.3299722752125247

-----SVR()-----

MAE: 3.3461896091157075

MSE: 355.3495094678607

RMSE: 18.85071641789406

R2 Square 0.2936190204342203

-----DecisionTreeRegressor()-----

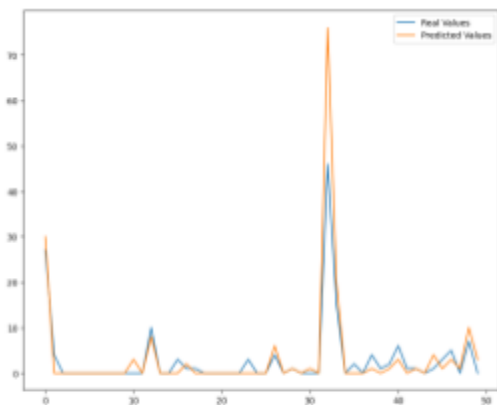
MAE: 2.838631907967744

MSE: 127.81410727276409

RMSE: 11.305490138546142

R2 Square 0.7459249221059455

Dari hasil pemodelan data dan evaluasi performa model pada data dengan skala Provinsi dapat dilihat apabila menggunakan metode Linier Regression memiliki hasil R2 sebesar 33%, sedangkan SVR memiliki hasil R2 sebesar 29%, dan Decision Tree Regressor memiliki hasil R2 sebesar 75%. Karena Decision Tree Regressor memiliki nilai R2 terbesar maka metode tersebut akan digunakan untuk memprediksi kasus kematian dengan kasus riil



Gambar 4. Prediksi kasus kematian dengan kasus riil menggunakan metode decision tree

Pada gambar 4 di atas masih terlihat adanya beberapa prediksi yang offset karena nilai R2 score yang "hanya" sekitar 75%. Hal itu terjadi karena gambar di atas membandingkan antara prediksi real time

dengan menggunakan Active cases covid 19 dengan tidak menggunakan Active cases covid 19.

4. SIMPULAN DAN SARAN

Kesimpulan yang dapat diambil dari penelitian ini adalah menemukan 5 Provinsi dengan jumlah kematian tertinggi akibat Covid-19 yang terjadi di Indonesia yaitu Jawa Tengah (33.4 ribu), Jawa Timur (31.7 ribu), Jawa Barat (15.9 ribu), DKI Jakarta (15.5 ribu), dan Daerah Istimewa Yogyakarta (5.9 ribu). Yang mana provinsi Jawa Tengah, Jawa Timur, Jawa Barat, dan DKI Jakarta juga termasuk 5 provinsi dengan jumlah kasus terbanyak apabila dilihat pada data jumlah kasus terjadinya Covid-19 di Indonesia.

Saran yang dapat diberikan kepada peneliti selanjutnya diharapkan melakukan pengkajian yang lebih dalam dan lebih terkini terkait jumlah kematian di Indonesia dan diharapkan bisa memakai jenis diagram yang lain sebagai upaya pengembangan Visualisasi data.

5. PUSTAKA ACUAN

Syafrida & Hartati, R.(2020). Bersama Melawan Virus Covid-19 di Indonesia. Jurnal Sosial dan Budaya Syar'i.

Wahyuni, E., Arifiyanti, A., Kustyani, M., (2019). Exploratory Data Analysis dalam Konteks Klasifikasi Data Mining. Jurnal Rekayasa Teknologi Industri dan Informasi (RTII).

<https://laporcovid19.org/>

<https://databoks.katadata.co.id/>