

Analisis Penyebaran COVID-19 di Sulawesi Selatan Menggunakan *Machine Learning* (Regression Models)

Nur Qadri Bahar^{1, a)} dan Nurul Maqfirah Rauf^{2, b)}

¹Program Studi Matematika, Institut Teknologi Kalimantan

²Program Studi Ilmu Aktuaria, Institut Teknologi Kalimantan
Balikpapan

a) nur.bahar@lecturer.itk.ac.id

b) nurul.rauf@lecturer.itk.ac.id

Abstrak. COVID-19 merupakan bencana yang sangat mengganggu manusia secara global abad ini. Berbagai macam dampak buruk diakibatkan oleh pandemi COVID-19, mulai dari dampak sosial, ekonomi, psikologi dan ekonomi. Dunia telah banyak belajar dari pandemi COVID-19, termasuk bagaimana mengatasi dan mengurangi dampak buruk akibat COVID-19 dengan berbagai macam ilmu dan perangkat, termasuk menggunakan *machine learning* dalam menganalisis bencana dan dampaknya. *Machine learning* menggunakan berbagai teknik dalam penerapannya, salah satunya adalah regresi yang mampu memberikan solusi bagi banyak permasalahan, termasuk menganalisis penyebaran COVID-19. Dengan *framework* yang dimiliki oleh bahasa pemrograman *python*, *machine learning* mampu menganalisis dan memberikan *insight* yang tersembunyi dari dataset yang dikoleksi oleh Dinas Kesehatan Provinsi Sulawesi Selatan. Dalam penelitian ini, dilakukan analisis penyebaran COVID-19 menggunakan model regresi linear biasa, regresi ridge, dan regresi LASSO. Model regresi ridge mampu menjelaskan variabilitas dataset (variabel target) dengan tingkat yang paling baik dengan $R^2 = 73,61\%$. Namun dari ketiga model yang dibangun, semuanya memiliki performa yang kurang, dengan MAPE yang sangat tinggi. Model regresi biasa memiliki MAPE sebesar 2.5722572116 pada *training* dan 3.0803084681 pada *testing*. Model regresi ridge memiliki MAPE sebesar 2.8757061838 pada *training* dan 3.0572481631. Model regresi LASSO memiliki MAPE sebesar 2.6645211829 pada *training* dan 3.0572481631 pada *testing*. Dengan MAPE yang tinggi pada *training* dan *testing*, artinya semua model terindikasi *underfitting*, sehingga perlu ditindaklanjuti dengan membangun model yang lebih kompleks.

Kata Kunci: COVID-19, *machine learning*, regresi

1. PENDAHULUAN

Sejak kasus pertamanya di kota Wuhan pada Desember 2019, hingga tanggal 21 September 2021 tercatat 228.807.631 individu terinfeksi, 4.697.099 kematian akibat virus corona, dan untuk kasus tertinggi dilaporkan terjadi di benua Amerika yakni sebanyak 88.207.746 kasus, dimana benua Asia berada di urutan ketiga sebanyak 42.594.207 kasus, sebanyak 4.195.958 kasus berasal dari Indonesia dengan jumlah kematian sebanyak 140.805 [1]. Dampak dari COVID-19 masih terasa bahkan setelah pandemi berlalu, mulai dari dampaknya terhadap ekonomi, kehidupan sosial, bahkan sampai kepada perubahan perilaku manusia. Oleh karena itu, penelitian terhadap COVID-19 menjadi penting untuk dilakukan. Sejak awal kemunculannya, penyebaran COVID-19 telah banyak diteliti oleh para ahli, dan telah banyak model prediksi yang dihasilkan mulai dari pengembangan model matematika SIR sampai kepada penerapan *machine learning* sebagai salah satu upaya dalam penanggulangan kasus penyebaran COVID-19. Di Indonesia, Provinsi Sulawesi Selatan menjadi salah satu provinsi dengan peningkatan kasus COVID-19 yang tinggi. Penelitian terdahulu telah banyak membahas perilaku penyebaran COVID-19 di Sulawesi Selatan dengan berbagai macam

metode, namun belum ada yang menggunakan *machine learning* khususnya *regression models* untuk menganalisis dan memprediksi perilaku penyebaran COVID-19 di Provinsi Sulawesi Selatan. Oleh karena itu, penting untuk melakukan analisis terkait penyebaran COVID-19 di Sulawesi Selatan dengan mengimplementasikan *machine learning*.

2. TINJAUAN PUSTAKA

1. COVID-19

Coronavirus Disease 2019 (COVID-19) adalah penyakit menular yang disebabkan oleh *Severe Acute Respiratory Syndrome Coronavirus 2* (SARS-CoV-2). SARS-CoV-2 merupakan *coronavirus* jenis baru yang belum pernah diidentifikasi sebelumnya pada manusia. Tanda dan gejala umum infeksi COVID-19 antara lain gejala gangguan pernapasan akut seperti demam, batuk, dan sesak napas. Masa inkubasi rata-rata 5-6 hari dengan masa inkubasi terpanjang 14 hari. Pada kasus COVID-19 yang berat dapat menyebabkan pneumonia, sindrom pernapasan akut, gagal ginjal dan bahkan kematian [1].

Terdapat empat kelompok yang dapat digunakan untuk membuat model penyebaran COVID-19, yaitu:

Kelompok Sehat/Rentan (selanjutnya disebut *Susceptible* dan disimbolkan $S(t)$), yaitu jumlah individu yang rentan atau berpeluang terinfeksi COVID-19 pada waktu t .

Kelompok Suspek (selanjutnya disebut *Exposed/Suspected* dan disimbolkan dengan $E(t)$), yaitu jumlah individu yang memiliki gejala infeksi saluran pernapasan akut atau riwayat kontak erat dengan individu terinfeksi COVID-19 pada waktu t .

Kelompok Dirawat/Terinfeksi (selanjutnya disebut *Infected* dan disimbolkan dengan $I(t)$), yaitu jumlah individu terinfeksi/dirawat karena COVID-19 pada waktu t .

Kelompok Sembuh (selanjutnya disebut *Recovered* dan disimbolkan dengan $R(t)$), yaitu jumlah individu sembuh dari COVID-19 pada waktu t [2].

2. Machine Learning

Machine Learning (ML) merupakan salah satu bidang dari kecerdasan buatan yang terus berkembang secara pesat. *Machine Learning* atau dalam bahasa Indonesia disebut pembelajaran mesin makin dibutuhkan dan kemampuannya kian meningkat disebabkan karena makin banyaknya perolehan data [6]. *Machine learning* merupakan bidang studi yang fokus kepada desain dan analisis algoritma sehingga memungkinkan komputer untuk dapat belajar. *Machine learning* juga dapat diartikan sebuah komputer yang memiliki kemampuan belajar tanpa deprogram secara eksplisit. Program tersebut memanfaatkan data untuk membangun model dan mengambil keputusan berdasarkan model yang telah dibangun [4].

Secara garis besar, definisi ML adalah bagaimana cara membuat program komputer yang meningkat berdasarkan pengalaman [5]. Definisi lain *machine learning* yaitu bidang kecerdasan buatan untuk memperoleh pengetahuan secara otomatis untuk menciptakan suatu sistem yang cerdas. Algoritma ML belajar berbasis pada input yang disediakan sehingga keberhasilan dalam algoritma ini pada ketersediaan dan kompleksitas data [2].

Dalam konsep *machine learning*, terdapat dua pendekatan yaitu *Supervised learning* dan *Unsupervised learning*.

2.1 Supervised Learning

Supervised learning merupakan metode pembelajaran mesin yang belajar dari data pelatihan berlabel untuk membantu pengguna dalam memprediksi hasil untuk data yang tidak terduga. Mesin dilatih menggunakan data berlabel dengan hasil prediksi yang benar [9]. Secara sederhana, *supervised learning* adalah metode machine learning di mana mesin diberikan serangkaian data lengkap yang pada umumnya kompleks dan besar agar mesin bisa mempelajari pola yang tersembunyi dari data.

Supervised learning dibagi menjadi dua yaitu regresi dan klasifikasi. Beberapa model yang tergolong dalam *supervised learning* untuk regresi adalah *Ordinary Linear Regression*, *Ridge Regression*, *Lasso Regression*, dan *Random Forest Regression*. Adapun untuk model yang tergolong dalam *unsupervised learning* untuk klasifikasi salah satunya adalah *K-Nearest Neighbor* (KNN).

2.2 Unsupervised Learning

Metode ML yang tidak memerlukan pengawasan model oleh pengguna merupakan definisi dari *unsupervised learning*. *Unsupervised learning* memungkinkan model menemukan pola dan informasi secara mandiri. *Unsupervised learning* berkaitan dengan data yang tidak berlabel [3]. Ketika hanya data input (X) yang tersedia dan tidak ada output yang cocok untuk variabel tersebut, maka berada dalam domain *unsupervised learning*. Tujuan dari jenis pembelajaran ini adalah untuk memodelkan struktur atau distribusi dasar data dengan maksud untuk mendapatkan pemahaman lebih lanjut tentang data tersebut. Masalah dalam *unsupervised learning* pengelompokan dan hubungan [5].

Unsupervised learning yang digunakan untuk mengungkapkan pola dalam data yang tidak memiliki label dikenal dengan istilah pengelompokan (*clustering*). Ini adalah jenis *machine learning* dimana tidak ada target yang telah ditentukan, melainkan sistem berusaha untuk mengelompokkan set data yang diberikan berdasarkan kesamaan pola yang ada di dalamnya. Metode ini tidak bergantung pada label atau keluaran yang telah ditetapkan untuk setiap data yang sedang dianalisis [10]. Pendekatan *unsupervised learning* menjadi bermanfaat ketika mencoba mengatasi masalah yang rumit.

3. Regresi

Regresi adalah salah satu teknik dalam machine learning, dimana teknik ini berasal dari ilmu statistik untuk mencari parameter-parameter dalam persamaan linear yang dapat memetakan input dan output. Regresi linear merupakan teknik tertua, paling mudah dan banyak digunakan pada supervised machine learning untuk memprediksi sebuah nilai atau keadaan. Teori ini merupakan salah satu teori paling penting abad ini. Tanpa teori regresi, ilmu tentang deep learning tidak akan ada.

Analisis regresi setidaknya memiliki tiga kegunaan, yaitu mendeskripsi fenomena data atau kasus yang sedang diteliti; untuk tujuan kontrol; serta untuk tujuan prediksi. Regresi mampu memprediksikan fenomena data melalui terbentuknya suatu model hubungan yang bersifat numerik. Inti dari regresi linear adalah memprediksi variabel target dengan mencocokkan dengan fungsi relasi linear terbaik di antara variabel bebas dan terikat. Regresi linear ada beberapa macam di antaranya regresi linear sederhana, regresi linear berganda, regresi logistik, regresi ridge, dan regresi lasso [4].

3. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah CRISP-DM (*Cross Industry Standard Process for Data Mining*). CRISP-DM menyediakan standar proses data mining sebagai strategi pemecahan masalah secara umum dari bisnis atau objek yang sedang diteliti.

1. Business Understanding

Tahap awal ini untuk memahami tujuan dan kebutuhan dari proyek secara detail dari sudut pandang bisnis atau objek penelitian. Domain knowledge atau pengetahuan secara menyeluruh terhadap objek penelitian akan memudahkan dalam menetapkan struktur dan alur penelitian sampai kepada penarikan *insight* dan kesimpulan.

2. Data Understanding

Tahap ini dimulai dengan mengumpulkan data yang akan dianalisis untuk mengenal lebih lanjut deskripsi dari data, mulai dari identifikasi awal untuk mendeteksi bagian-bagian menarik dan potensial dari data yang kemudian bisa digunakan untuk hipotesis awal.

3. Data Preparation

Tahap ini adalah kegiatan untuk membangun dataset, proses awal terhadap data mentah sehingga diperoleh dataset akhir yang siap untuk dimodelkan. Tahap ini bisa dilakukan berulang untuk memperoleh dataset yang bersih dan kaya akan informasi.

4. Modelling

Tahap ini akan dipilih dan diterapkan beberapa algoritma dan teknik pemodelan sesuai dengan jenis data dan tujuan yang hendak dicapai.

5. Evaluation

Pada tahap ini dilakukan evaluasi terhadap performa model berdasarkan asumsi umum dalam *machine learning* maupun asumsi yang khusus berkaitan dengan model tertentu yang digunakan.

6. Deployment

Pada tahap ini, pengetahuan atau informasi yang telah diperoleh akan diatur dan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh pengguna [8].

3. HASIL DAN PEMBAHASAN

Implementasi analisis data pada penelitian ini terdiri dari beberapa tahapan yang didasarkan pada konsep metodologi CRISP-DM.

A. Pemahaman Bisnis (Business Understanding)

Pada tahap ini diperlukan pemahaman tentang latar belakang dan tujuan pada proses bisnis yang berhubungan dengan penyebaran COVID-19 di Provinsi Sulawesi Selatan, di antaranya:

1. Menentukan Tujuan Bisnis
2. Menilai Situasi
3. Menentukan Tujuan Analisis Data

B. Pemahaman Data (Data Understanding)

Tahapan pemahaman data diawali dengan pengumpulan data awal atau data mentah dan akan melakukan pengolahan data, hasil kegiatan ini dalam rangka untuk membiasakan diri dengan data guna untuk membantu mengidentifikasi masalah data.

1. Mengumpulkan Data Awal (Collect the Initial Data)

Pengumpulan data dilakukan dengan mengajukan permintaan dataset COVID-19 yang ada di Dinas Kesehatan Provinsi Sulawesi Selatan.

2. Pemeriksaan Data (Editing)

Pemeriksaan data dilakukan sebagai langkah awal untuk mengurangi dataset yang tidak valid dan menghapus atribut yang tidak dibutuhkan untuk penelitian ini. Dari proses ini didapatkan dataset dalam format csv dari tanggal 19 Maret 2020 sampai dengan 31 Desember 2021 yang berjumlah 653 data yang terdiri dari delapan atribut. Sampel dataset dapat dilihat pada Gambar 1.

| Tanggal | Suspect (S) | Exposed (E) | Infected (I) | Recovered (R) | Meninggal | Total kasus | N | 05-Dec-21 | 9059997 | 0 | 0 | 3 | 0 | 3 | 9.060.000 |
|-----------|-------------|-------------|--------------|---------------|-----------|-------------|-----------|-----------|---------|---|----|---|---|----|-----------|
| 19-Mar-20 | 9059998 | 0 | 2 | 0 | 0 | 2 | 9.060.000 | 06-Dec-21 | 9059995 | 0 | 3 | 2 | 0 | 5 | 9.060.000 |
| 20-Mar-20 | 9059930 | 70 | 0 | 0 | 1 | 71 | 9.060.000 | 07-Dec-21 | 9059995 | 0 | 1 | 4 | 0 | 5 | 9.060.000 |
| 21-Mar-20 | 9059982 | 18 | 0 | 0 | 0 | 18 | 9.060.000 | 08-Dec-21 | 9059995 | 0 | 3 | 2 | 1 | 6 | 9.060.000 |
| 22-Mar-20 | 9059955 | 45 | 0 | 0 | 0 | 45 | 9.060.000 | 09-Dec-21 | 9059994 | 0 | 3 | 3 | 0 | 6 | 9.060.000 |
| 23-Mar-20 | 9059941 | 59 | 0 | 0 | 2 | 61 | 9.060.000 | 10-Dec-21 | 9059995 | 0 | 0 | 5 | 0 | 5 | 9.060.000 |
| 24-Mar-20 | 9059977 | 21 | 2 | 0 | 0 | 23 | 9.060.000 | 11-Dec-21 | 9059984 | 0 | 10 | 6 | 0 | 16 | 9.060.000 |
| 25-Mar-20 | 9059957 | 34 | 9 | 0 | 1 | 44 | 9.060.000 | 12-Dec-21 | 9059994 | 0 | 2 | 4 | 0 | 6 | 9.060.000 |
| 26-Mar-20 | 9059926 | 60 | 14 | 0 | 0 | 74 | 9.060.000 | 13-Dec-21 | 9059997 | 0 | 0 | 3 | 0 | 3 | 9.060.000 |
| 27-Mar-20 | 9059887 | 111 | 2 | 0 | 0 | 113 | 9.060.000 | 14-Dec-21 | 9059994 | 0 | 5 | 1 | 0 | 6 | 9.060.000 |
| 28-Mar-20 | 9059970 | 27 | 3 | 0 | 0 | 30 | 9.060.000 | 15-Dec-21 | 9059994 | 0 | 3 | 3 | 0 | 6 | 9.060.000 |
| 29-Mar-20 | 9059843 | 142 | 15 | 0 | 1 | 158 | 9.060.000 | 16-Dec-21 | 9059993 | 0 | 3 | 4 | 0 | 7 | 9.060.000 |
| 30-Mar-20 | 9059859 | 137 | 4 | 0 | 1 | 142 | 9.060.000 | 17-Dec-21 | 9059997 | 0 | 3 | 0 | 0 | 3 | 9.060.000 |
| 31-Mar-20 | 9059930 | 70 | 0 | 0 | 0 | 70 | 9.060.000 | 18-Dec-21 | 9059998 | 0 | 0 | 2 | 0 | 2 | 9.060.000 |
| 01-Apr-20 | 9059601 | 385 | 14 | 0 | 0 | 399 | 9.060.000 | 19-Dec-21 | 9059991 | 0 | 4 | 5 | 0 | 9 | 9.060.000 |
| 02-Apr-20 | 9059755 | 245 | 0 | 0 | 2 | 247 | 9.060.000 | 20-Dec-21 | 9059995 | 0 | 2 | 3 | 0 | 5 | 9.060.000 |
| 03-Apr-20 | 9059522 | 463 | 15 | 0 | 1 | 479 | 9.060.000 | 21-Dec-21 | 9059997 | 0 | 2 | 1 | 0 | 3 | 9.060.000 |
| 04-Apr-20 | 9059591 | 405 | 0 | 4 | 2 | 411 | 9.060.000 | 22-Dec-21 | 9059990 | 0 | 5 | 5 | 1 | 11 | 9.060.000 |
| 05-Apr-20 | 9059865 | 124 | 2 | 9 | 0 | 135 | 9.060.000 | 23-Dec-21 | 9059992 | 0 | 3 | 5 | 0 | 8 | 9.060.000 |
| 06-Apr-20 | 9059802 | 163 | 30 | 5 | 1 | 199 | 9.060.000 | 24-Dec-21 | 9059993 | 0 | 3 | 4 | 0 | 7 | 9.060.000 |
| 07-Apr-20 | 9059874 | 115 | 10 | 1 | 2 | 128 | 9.060.000 | 25-Dec-21 | 9059996 | 0 | 3 | 1 | 0 | 4 | 9.060.000 |
| 08-Apr-20 | 9059946 | 47 | 5 | 2 | 1 | 55 | 9.060.000 | 26-Dec-21 | 9059999 | 0 | 0 | 1 | 0 | 1 | 9.060.000 |
| 09-Apr-20 | 9059909 | 83 | 8 | 0 | 0 | 91 | 9.060.000 | 27-Dec-21 | 9059997 | 0 | 0 | 3 | 0 | 3 | 9.060.000 |
| 10-Apr-20 | 9059886 | 81 | 33 | 0 | 0 | 114 | 9.060.000 | 28-Dec-21 | 9059987 | 0 | 10 | 3 | 0 | 13 | 9.060.000 |
| 11-Apr-20 | 9059879 | 106 | 11 | 4 | 2 | 123 | 9.060.000 | 29-Dec-21 | 9059999 | 0 | 1 | 0 | 0 | 1 | 9.060.000 |
| | | | | | | | | 30-Dec-21 | 9059995 | 0 | 4 | 1 | 0 | 5 | 9.060.000 |
| | | | | | | | | 31-Dec-21 | 9059995 | 0 | 1 | 4 | 0 | 5 | 9.060.000 |

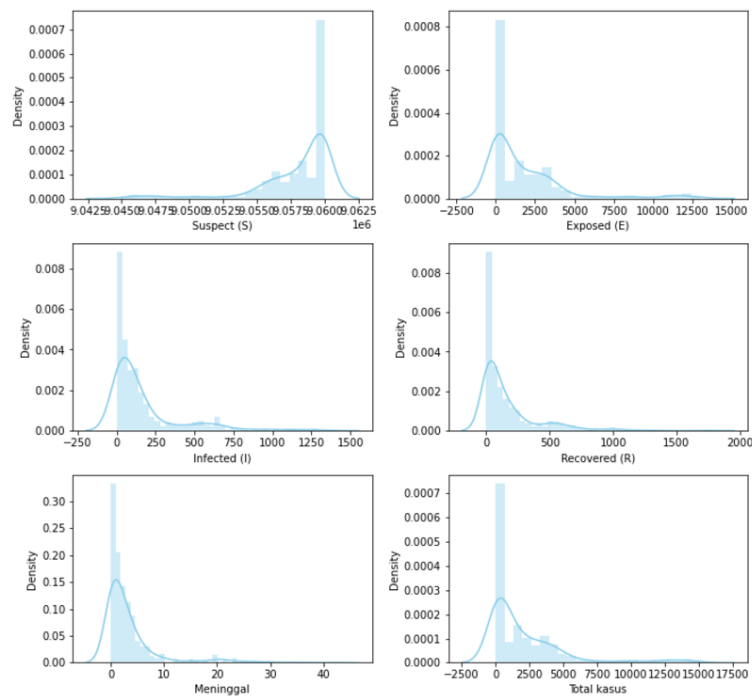
Gambar 1. Dataset Mentah Kasus COVID-19 di Provinsi Sulawesi Selatan Maret 2020-Desember 2021

C. Persiapan Data (Data Preparation)

Data yang digunakan akan diolah dengan menggunakan bahasa pemrograman *python*. Tahapan awal adalah dengan melakukan analisis data secara deskriptif terhadap dataset dengan atribut yang telah dipilih dari proses *editing* sebelumnya.. Dari proses analisis data deskriptif diketahui bahwa sebagian besar atribut berdistribusi tidak normal.

| | Suspect (S) | Exposed (E) | Infected (I) | Recovered (R) | Meninggal | Total kasus | N |
|--------------|--------------|--------------|--------------|---------------|------------|--------------|-----------|
| count | 6.530000e+02 | 653.000000 | 653.000000 | 653.000000 | 653.000000 | 653.000000 | 653.0 |
| mean | 9.057866e+06 | 1800.736600 | 168.453292 | 164.966309 | 3.431853 | 2137.588055 | 9060000.0 |
| std | 3.005636e+03 | 2614.986846 | 235.060441 | 232.296930 | 5.644941 | 3010.516262 | 0.0 |
| min | 9.044860e+06 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 1.000000 | 9060000.0 |
| 25% | 9.056879e+06 | 69.000000 | 27.000000 | 14.000000 | 0.000000 | 201.000000 | 9060000.0 |
| 50% | 9.059247e+06 | 610.000000 | 83.000000 | 74.000000 | 2.000000 | 753.000000 | 9060000.0 |
| 75% | 9.059800e+06 | 2747.000000 | 180.000000 | 206.000000 | 4.000000 | 3126.000000 | 9060000.0 |
| max | 9.059999e+06 | 13022.000000 | 1359.000000 | 1763.000000 | 42.000000 | 15170.000000 | 9060000.0 |

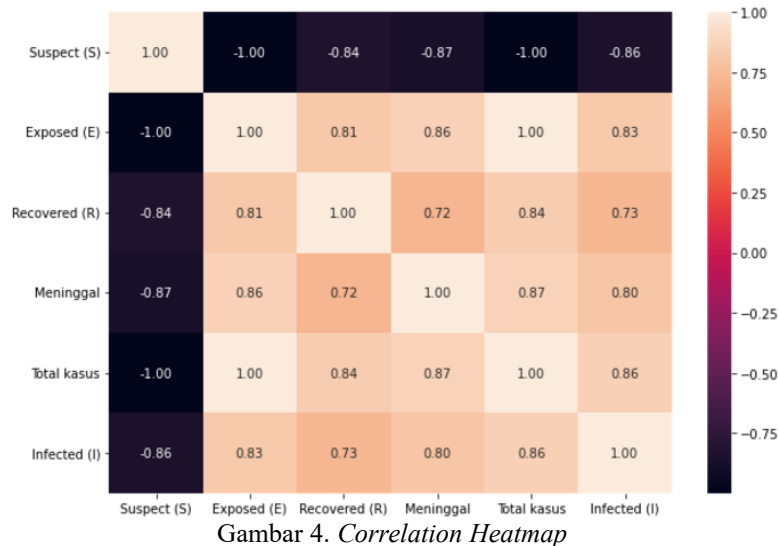
Gambar 2. Deskripsi Sebaran Data



Gambar 3. Grafik Distribusi Data

D. Pemodelan dan Evaluasi (Modelling and Evaluation)

Pada tahap *modelling* akan dibangun tiga model regresi, yaitu regresi linear biasa, regresi ridge, dan regresi LASSO. Tahap awal dimulai dengan *splitting data* menjadi data *train* dan *test*, dilanjutkan dengan studi multikolinieritas dengan mengimplementasikan metode *VIF score* dan *correlation heatmap*.



Gambar 4. *Correlation Heatmap*

Dari hasil studi multikolinearitas, Gambar 4 menunjukkan bahwa hanya atribut “Recovered (R)” yang tidak terindikasi mengalami multikolinearitas. Sedangkan variabel independen atau atribut lainnya terindikasi mengalami multikolinearitas dengan tingkat korelasi antar variabel di atas 0.8. Beberapa atribut dihapus (*drop*) sehingga tersisa tiga atribut saja yaitu “Infected (I)” sebagai variabel dependen (*target variable*), “Suspect (S)” dan “Recovered (R)” sebagai variabel independen (*fitures*).

Selanjutnya dilakukan *training* data menggunakan model regresi linear biasa, didapatkan bentuk umum model regresi linear biasa $Y = -0.005212 - 0.823228X_1 + 0.034946X_2$ di mana X_1 adalah “Suspect (S)” dan X_2 adalah “Recovered (R)”. Hasil dari *training* model regresi linear biasa ini bisa memodelkan variabilitas dari target variabel dengan $R^2 = 0.7315$.

Selanjutnya, dibangun model regresi dengan regularisasi yaitu regresi ridge dan regresi LASSO sebagai pembanding. Kedua model ini memiliki hasil yang cenderung sama dengan model regresi linear biasa. Dengan menggunakan metrik pengukuran yang sama (MAPE), model regresi ridge dan regresi LASSO memiliki performa seperti ditunjukkan pada Tabel 1.

Tabel 1. Perbandingan Performa Model Regresi

| Model | MAPE | | R^2 |
|----------------------|--------------------|--------------------|--------------------|
| | <i>Training</i> | <i>Testing</i> | |
| Regresi Linear Biasa | 2.5722572115891102 | 3.0803084681401507 | 0.7315249792790186 |
| Regresi Ridge | 2.8757061838436290 | 3.0572481631003274 | 0.7360831193030035 |
| Regresi LASSO | 2.6645211829457818 | 3.0572481631003274 | 0.7278150479811454 |

Dari Tabel 1 diketahui bahwa semua model mampu membaca variabilitas dari variabel target dengan baik. Namun, ketiga model yang dibangun memiliki performa yang kurang dengan nilai MAPE yang sangat tinggi baik pada *training* maupun pada *testing*. Artinya, semua model yang dibangun terindikasi mengalami *underfitting* sehingga belum bisa untuk dilanjutkan ke tahap *deployment*.

4. KESIMPULAN

Berdasarkan analisis dan pengujian pada dataset COVID-19 di Sulawesi Selatan, diketahui bahwa ketiga algoritma regresi yang digunakan memiliki MAPE yang tinggi, sehingga tidak digunakan oleh para ahli dan *stake holder* di bidang kesehatan Provinsi Sulawesi Selatan dalam memprediksi COVID-19 dan bagi seorang peneliti dapat dijadikan bahan untuk penelitian selanjutnya menggunakan algoritma *machine learning* yang lebih kompleks seperti XGBoost dan Random Forest.

REFERENSI

- [1] Bahar, N.Q. (2022). *Model Matematika SEIRS pada Penyebaran COVID-19 dengan Vaksinasi dan Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) di Provinsi Sulawesi Selatan*. (Unpublished magister thesis). Universitas Negeri Makassar, Makassar.
- [2] Hartati, S. (2021). *Kecerdasan Buatan Berbasis Pengetahuan*. Yogyakarta: UGM Press.
- [3] Hastie, T., Tibshirani, R., Friedman, J. (2009). *Unsupervised Learning*. In: *The Elements of Statistical Learning*. Springer Series in Statistic. New York, NY: Springer.
- [4] Id, I.D. (2021). *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python*.
- [5] Karim, M. A., & Yulida, Y. (2021). Analisis Kestabilan dan Sensitivitas pada Model Matematika SEIRD dari Penyebaran COVID-19: Studi Kasus di Kalimantan Selatan. *Journal Binawakya*, 16, hlm 7003-7012.
- [6] Mitchell, T.M. (1997). *Machine Learning*. New York: McGraw-Hill.
- [7] Muflikhah, L., Mahmudy, W.F., & Kurnianingtyas, D. (2023). *Machine Learning*. Dikutip dari <http://books.google.co.id>.
- [8] Murti, E.W. (2024). Analisis dan Perbandingan Algoritma Prediksi dalam Mengetahui Perkiraan Peningkatan Jumlah Kasus COVID-19 di Kabupaten Boyolali dengan Metodologi CRISP-DM. *Jikes: Jurnal Ilmu Kesehatan*, 3, hlm 24-34.
- [9] Permana A.A, Wahyuddin, S., Santos, L.W., et al. (2023). *Machine Learning*. Padang: Global Eksekutif Teknologi.
- [10] Rokach, L. & Maimon, O. (2005). *Clustering Methods*. In: Maimon, O., Rokach, L. (eds). *Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer.