

Analisis Tingkat Kerawanan COVID-19 Menggunakan Metode *Random Forest* (Studi Kasus: Unit Pengembangan II Kertajaya dan III Tambak Wedi Kota Surabaya)

Analysis of the COVID-19 Susceptibility Level Using the Random Forest Method (Case Study: Development Unit II Kertajaya and III Tambak Wedi, Surabaya City)

Andy Kusuma Deradjat, Hepi Hapsari Handayani*

Departemen Teknik Geomatika, FTSLK-ITS, Kampus ITS Sukolilo, Surabaya, 60111, Indonesia

*Korespondensi Penulis: hapsari@geodesi.its.ac.id

Diterima: 31072022; Diperbaiki: 04062024; Disetujui: 09062024; Dipublikasi: 19062024

Abstrak: Pada bulan Maret 2020 pandemi COVID-19 telah melanda seluruh dunia termasuk Indonesia. Oleh karena itu, dinas kesehatan, pemerintah, dan layanan publik perlu bekerja sama secara global untuk mencegah penyebaran lebih lanjut. Analisis dan pemetaan kerawanan sangat dibutuhkan untuk mengembangkan informasi terkait pandemi COVID-19 dari beberapa faktor yang mempengaruhi penularan COVID-19, salah satunya adalah faktor sosial-ekonomi yang memiliki tingkat interaksi antar manusia yang cukup beresiko. Dengan *Machine Learning Technique Random forest* dapat membantu dalam memberikan keputusan tingkatan atau prioritas parameter dalam analisis dan pemetaan kerawanan pandemi COVID-19 dengan akurat dan mendukung keputusan serta tindakan pencegahan. Penelitian ini bertujuan untuk melakukan analisis dan pemetaan kerawanan pandemi COVID-19 berdasarkan faktor sosial-ekonomi di Unit Pengembangan II Kertajaya dan III Tambak Wedi Kota Surabaya. Penelitian ini menggunakan data statistik jumlah kasus penderita COVID-19 pada bulan Maret tahun 2020 hingga Agustus tahun 2021 dan data fasilitas sosial-ekonomi berupa jumlah fasilitas kesehatan (rumah sakit, puskesmas, dan klinik), objek wisata (tempat wisata dan taman kota), perbankan (atm dan bank), pasar, mall, SPBU, dan terminal. Hasil pengolahan random forest didapatkan tingkat pengaruh dari setiap parameter terhadap penyebaran COVID-19, yaitu fasilitas kesehatan 24,135%, terminal 20,338%, objek wisata 19,916%, mall 19,574%, pasar 11,317%, perbankan 2,628%, dan SPBU 2,092%. Berdasarkan penelitian terdahulu, uji akurasi model yang dihasilkan dapat dikatakan baik dengan nilai akurasi sebesar 0,946 dan nilai kappa sebesar 0,892 serta nilai AUC sebesar 0,984. Hasil pemetaan dari model random forest tersebut, didapatkan daerah dengan tingkat kerawanan rendah pada bagian sebelah utara Kecamatan Kenjeran dan sebelah timur Kecamatan Sukolilo dan Kecamatan Mulyorejo, sedangkan tingkat kerawanan tinggi pada bagian tengah Kecamatan Mulyorejo, dan bagian tengah Kecamatan Sukolilo.

Copyright © 2024 Geoid. All rights reserved.

Abstract: In March 2020 the COVID-19 pandemic had hit the whole world, including Indonesia. Therefore, health services, governments and public services need to work together globally to prevent further spread. Analysis and mapping is urgently needed to develop information related to the COVID-19 pandemic from several factors that influence the transmission of COVID-19, one of which is socio-economic factors that have a fairly risky level of human interaction. With *Machine Learning Technique Random forest* can assist in making decisions on the level or priority of parameters in analysis and mapping the susceptibility of the COVID-19 pandemic accurately and support decisions and preventive actions. This study aims to conduct analysis and mapping of the susceptibility of the COVID-19 pandemic based on socio-economic factors in the Kertajaya Development Unit II and III Tambak Wedi Surabaya City. This study uses statistical data on the number of cases of COVID-19 sufferers from March 2020 to August 2021 and data on socio-economic facilities in the form of the number of health facilities (hospitals, health centers, and clinics), tourist attractions (tourist attractions and city parks), banking (atm and bank), market, mall, gas station, and terminal. The results of random forest processing obtained the level of influence of each parameter on the spread of COVID-19, namely health facilities 24.135%, terminals 20.338%, tourist attractions 19.916%, malls 19.574%, markets 11.317%, banking 2.628%, and gas stations 2.092%. Based on previous research, the accuracy of the resulting model can be said to be good with an accuracy value of 0.946 and a kappa value of 0.892 and an AUC value of 0.984. The mapping results from the random forest model, obtained areas with a low level of susceptibility in the northern part of Kenjeran Sub-District and the eastern part of Sukolilo Sub-District and Mulyorejo Sub-District, while the high susceptibility level is in the middle part of Mulyorejo Sub-District and the middle part of Sukolilo Sub-District.

Kata kunci: COVID-19, Random forest, Sosial-Ekonomi

Deradjat, A.K., & Handayani, H.H. (2024). Analisis Tingkat Kerawanan COVID-19 Menggunakan Metode *Random Forest* (Studi Kasus: Unit Pengembangan II Kertajaya dan III Tambak Wedi Kota Surabaya). *Geoid*, 19(2), 256 - 265.

Pendahuluan

COVID-19 adalah penyakit pernapasan yang disebabkan oleh virus corona RNA untai tunggal yang sangat menular, SARS-CoV-2 (Chen dkk, 2020). Hal ini pertama kali diamati di Wuhan, ibu kota provinsi Hubei di Republik Rakyat Tiongkok, pada Desember 2019 (Zhu dkk, 2020). Virus Corona adalah ancaman pandemi berkelanjutan yang telah menyebar ke seluruh dunia dengan tingkat kematian sejumlah 5.377.197 yang tercatat dan 277.161.199 infeksi global yang dikonfirmasi (per 22 Desember 2021). Pada bulan Maret 2020 pandemi COVID-19 telah masuk melanda Indonesia. Indonesia merupakan salah satu negara yang juga mengalami pandemi COVID-19 dengan 4.261.072 kasus infeksi yang dikonfirmasi dan jumlah kematian telah mencapai 144.034 (per 22 Desember 2021).

Oleh karena itu, dinas kesehatan, pemerintah, dan layanan publik perlu bekerja sama secara global untuk mencegah penyebaran lebih lanjut. Banyak faktor yang mempengaruhi penularan suatu virus, termasuk sosial-ekonomi, agro-ekologi, dan meteorologi. Setiap penyakit dan pemetaan kerawanan yang terkait dengannya memiliki beberapa faktor yang efektif, dan setiap penyakit dapat dikaitkan dengan iklim, air, hewan, manusia, dan bahkan tanah (Brevik dkk, 2019; Kumar dkk, 2019; Yadav dkk, 2020). Penyebaran penyakit pandemi juga sangat dipengaruhi oleh aktivitas perkotaan yang mengidentifikasi hubungan antara penggunaan lahan perkotaan dan penularan COVID-19 (Brizuela dkk, 2021).

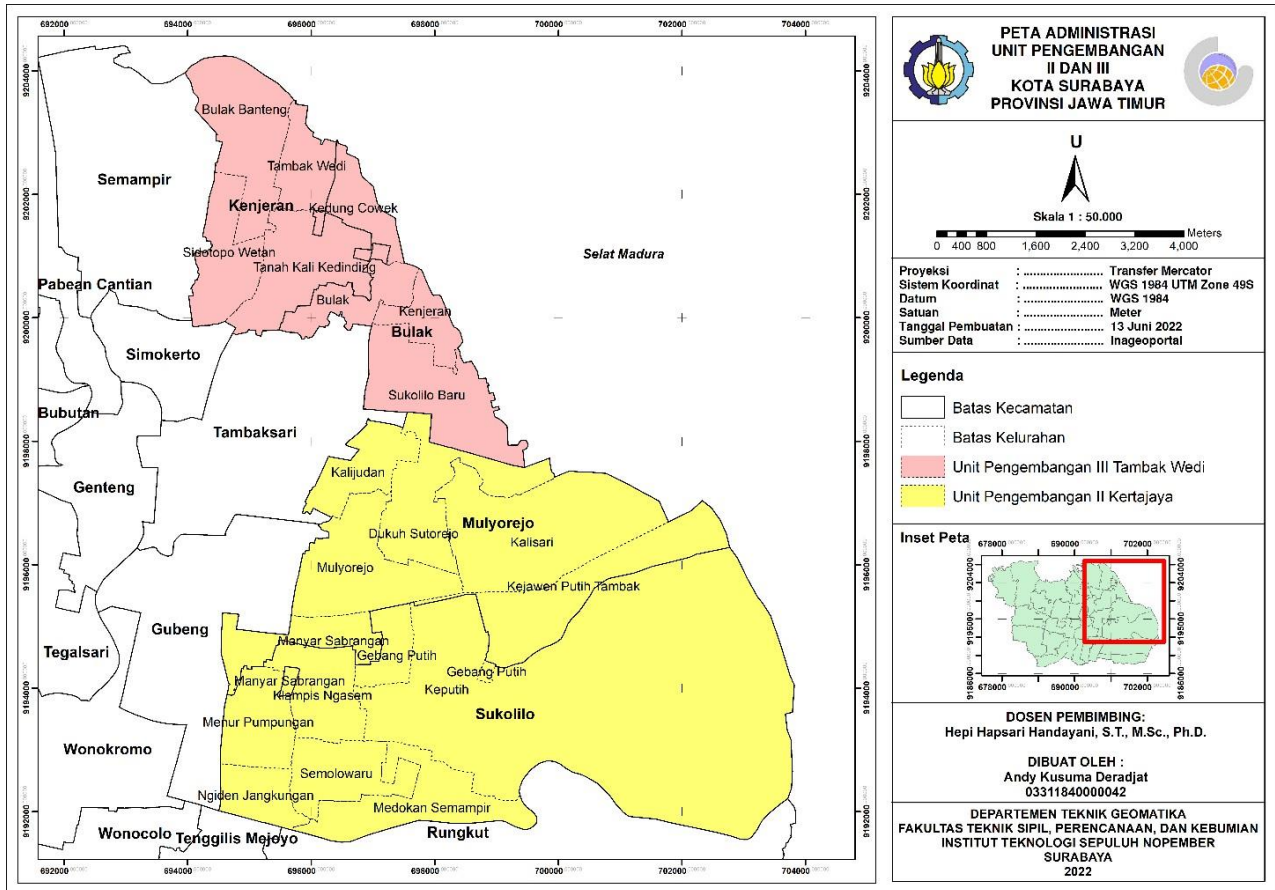
Berdasarkan hal tersebut, analisis dan pemetaan kerawanan sangat dibutuhkan untuk mengembangkan informasi terkait pandemi COVID-19 dari beberapa faktor yang mempengaruhi penularan COVID-19, salah satunya adalah faktor sosial-ekonomi yang meliputi beberapa parameter, antara lain objek wisata, fasilitas kesehatan, fasilitas niaga, dan fasilitas transportasi dikarenakan pada tempat tersebut sering dilakukannya interaksi antar manusia. Penting juga untuk mengembangkan informasi yang kuat dengan cepat menggunakan metode yang tidak bisa dan handal untuk memberikan kesadaran situasional dan untuk meningkatkan respon terhadap pandemi (Wu dkk, 2020). Random forest adalah salah satu jenis algoritma klasifikasi yang terdiri dari lebih satu pohon keputusan yang setiap pohon keputusan dibentuk bergantung pada nilai-nilai vektor acak sampel secara independen dan identik didistribusikan yang sama untuk semua pohon (Breiman, 2011).

Dengan *Machine Learning Technique Random forest* dapat membantu dalam memberikan keputusan tingkatan atau prioritas parameter dalam analisis dan pemetaan kerawanan pandemi COVID-19 dengan akurat dan mendukung keputusan serta tindakan pencegahan.

Pada penelitian ini bertujuan untuk melakukan analisis dan pemetaan kerawanan pandemi COVID-19 berdasarkan faktor sosial-ekonomi. Lokasi yang dipilih adalah Unit Pengembangan II Kertajaya dan III Tambak Wedi Kota Surabaya dan metode yang digunakan adalah *Machine Learning Technique Random forest*. Hasil dari penelitian ini diharapkan dapat menghasilkan investigasi komprehensif pertama yang dapat membantu pengelolaan dan pengendalian krisis COVID-19.

Data dan Metode

Lokasi penelitian yang digunakan adalah Unit Pengembangan II Kertajaya dan III Tambak Wedi Kota Surabaya, Provinsi Jawa Timur yang meliputi Kecamatan Bulak dengan jumlah kelurahan sebanyak 4 kelurahan, Kenjeran dengan jumlah kelurahan sebanyak 4 kelurahan, Mulyorejo dengan jumlah kelurahan sebanyak 6 kelurahan, dan Sukolilo dengan jumlah kelurahan sebanyak 7 kelurahan.



Gambar 1. Lokasi Penelitian

Data yang digunakan pada penelitian ini adalah Peta Batas Administrasi Unit Pengembangan II Kertajaya dan III Tambak Wedi yang meliputi Kecamatan Bulak, Kenjeran, Mulyorejo, dan Sukolilo, Kota Surabaya tahun 2020 dalam format *shapefile* dengan skala 1:5000, data statistik jumlah kasus penderita COVID-19 di setiap kelurahan yang berada di Unit Pengembangan II Kertajaya dan III Tambak Wedi pada bulan Maret tahun 2020 hingga Agustus tahun 2021, jumlah fasilitas pariwisata (tempat wisata dan taman), jumlah fasilitas kesehatan (rumah sakit, klinik, dan puskesmas), jumlah fasilitas perniagaan (mall, pasar, bank, dan ATM), dan jumlah fasilitas transportasi (terminal dan SPBU).

Proses pengolahan data diawali dengan tahap persiapan yang melingkupi identifikasi masalah dan studi literatur terkait analisis dan pemetaan kerawanan pandemi COVID-19 menggunakan *Machine Learning Technique Random Forest*. Selanjutnya dilakukan pengumpulan data fasilitas sosial-ekonomi dan data statistik jumlah kasus penderita COVID-19 pada bulan Maret tahun 2020 hingga Agustus tahun 2021.

Pengolahan fasilitas sosial-ekonomi dilakukan dengan menggunakan teknik *euclidean distance* pada perangkat lunak pengolahan data spasial untuk mendapatkan analisis radius jarak setiap parameter yang kemudian dilakukan *extract value* dengan data kejadian COVID-19.

Kemudian, dilakukan pembuatan data *training* sebanyak 70% dan data *testing* sebanyak 30% secara acak dengan menggunakan perangkat lunak R-Studio menggunakan *package caTools* yang nantinya akan digunakan untuk pengolahan dalam model *random forest* dengan menggunakan perangkat lunak R-Studio.

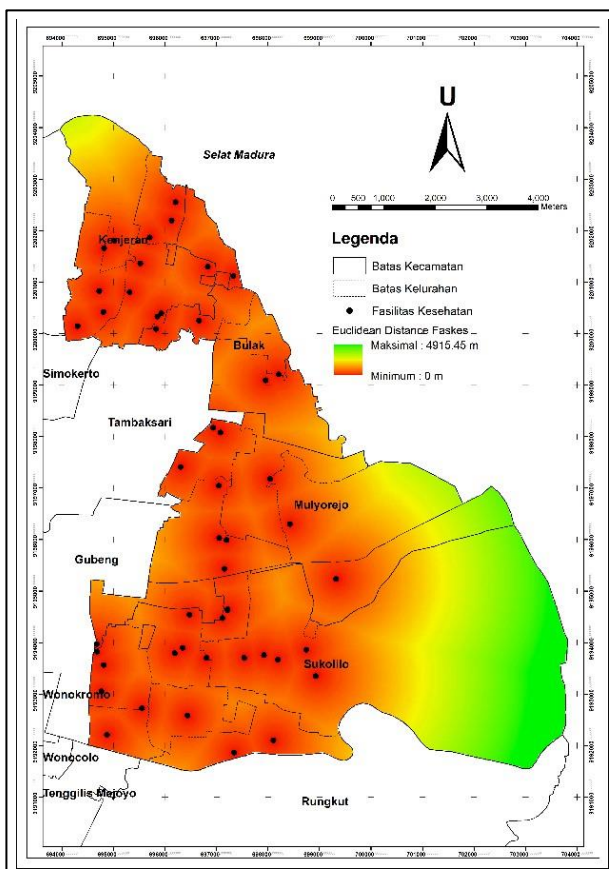
Melakukan uji akurasi pengolahan *random forest* menggunakan *confusion matrix* untuk mendapatkan nilai *accuracy*, *kappa*, *sensitivity*, *specificity*, dan kurva ROC untuk mendapatkan akurasi klasifikasi secara visual dengan ketentuan nilai AUC lebih besar dari 0,8 atau 80% dan nilai kappa lebih dari 0,6 atau 60%.

Selanjutnya, didapatkan *indeks importance degree* atau pembobotan pengaruh setiap parameter terhadap penyebaran pandemi COVID-19 yang akan digunakan dalam pembuatan peta kerawanan COVID-19. Pembuatan peta kerawanan COVID-19 dilakukan dengan cara pembobotan menggunakan *sum weighted tools* dari hasil *index importance degree* dan data raster pemodelan *euclidean distance* pada setiap parameter dengan menggunakan perangkat lunak pengolahan data spasial yang akan menghasilkan pemetaan kerawanan pandemi COVID-19 di Unit Pengembangan II Kertajaya dan III Tambak Wedi yang meliputi Kecamatan Bulak, Kenjeran, Mulyorejo, dan Sukolilo, Kota Surabaya.

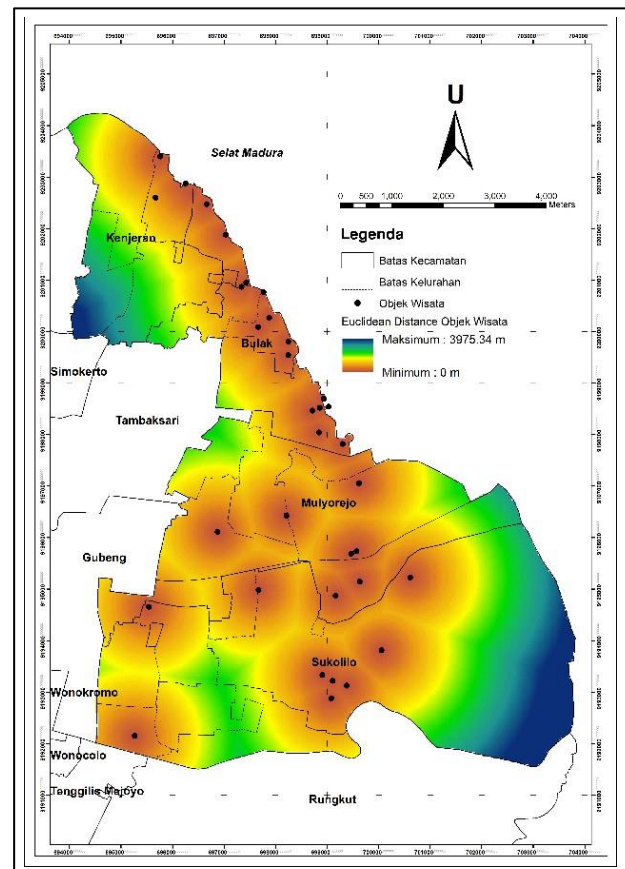
Hasil dan Pembahasan

1. Pemodelan Euclidean Distance

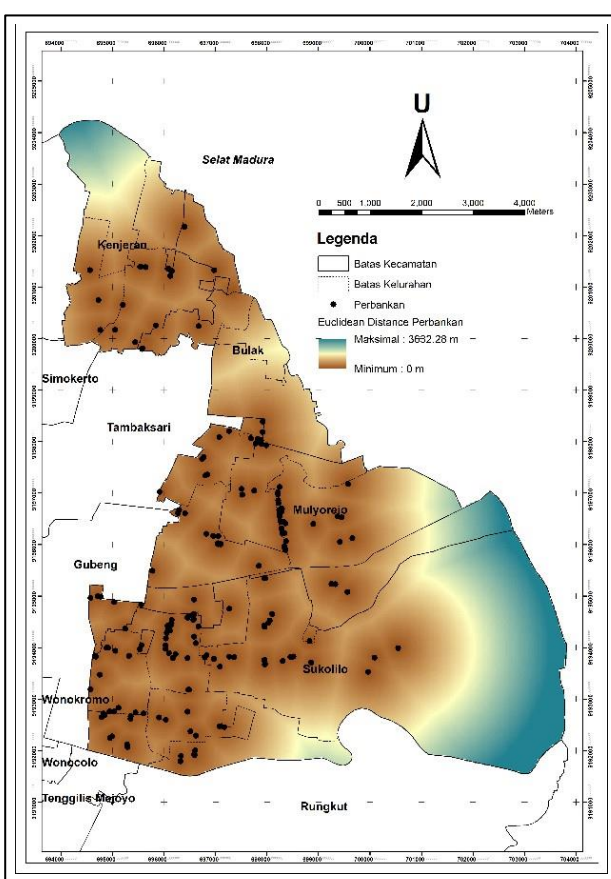
Penyebaran kasus COVID-19 tidak terlepas dari aktivitas sosial yang menyebabkan adanya interaksi maupun kontak fisik antar manusia. Pada awal masa pandemi COVID-19 kebijakan dalam menangani penyebaran COVID-19 dengan aktivitas sosial pada fasilitas umum masih dinamis, sehingga menyebabkan penyebaran COVID-19 meningkat cukup pesat. Dalam penelitian ini, dipilih 7 fasilitas sosial-ekonomi yang memiliki tingkat interaksi antar manusia yang cukup beresiko, antara lain fasilitas kesehatan (rumah sakit, puskesmas, dan klinik) yang berjumlah 49 objek, objek wisata (tempat wisata dan taman) yang berjumlah 34 objek, perbankan (atm dan bank) yang berjumlah 193 objek, pasar yang berjumlah 18 objek, mall yang berjumlah 3 objek, SPBU yang berjumlah 17 objek, dan terminal yang berjumlah 2 objek yang divisualisasikan seperti pada Tabel 1.



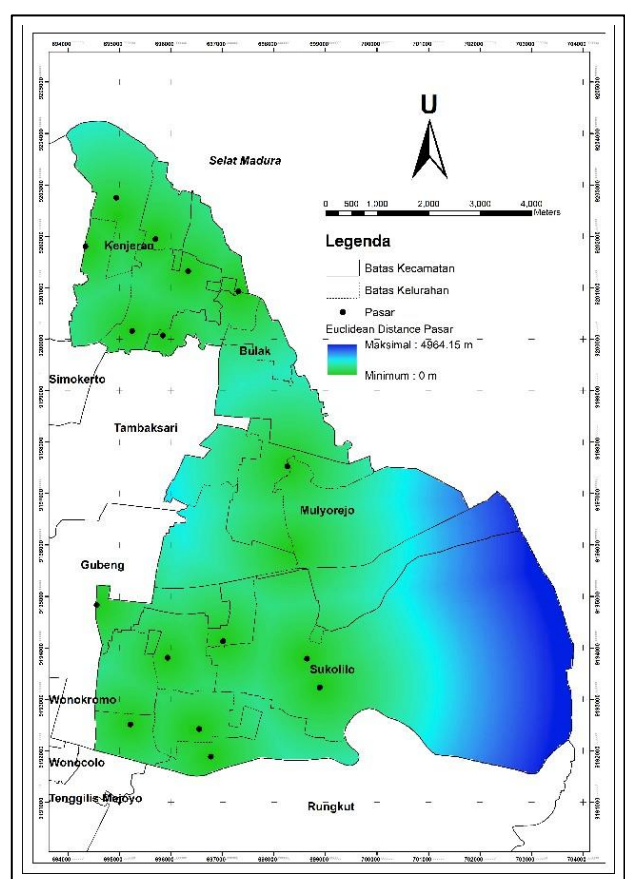
(a)



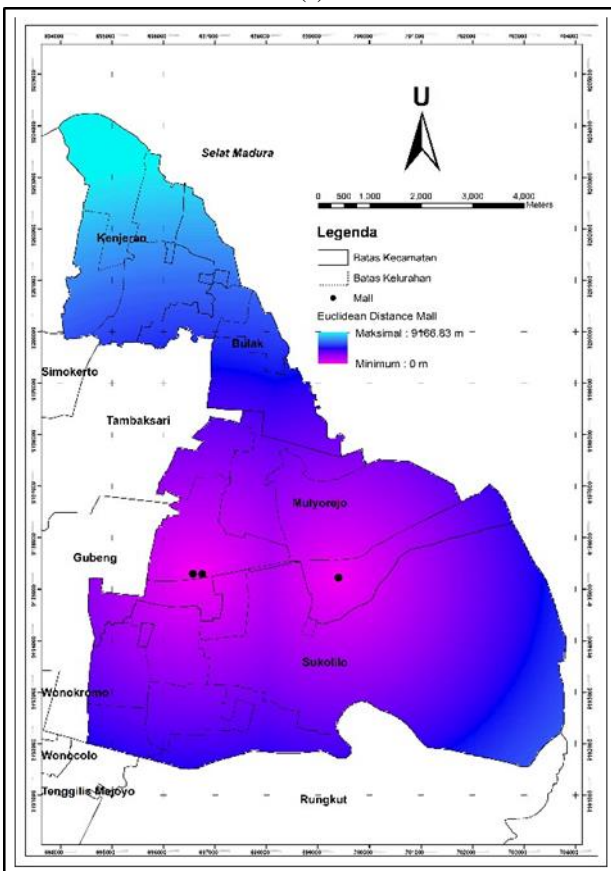
(b)



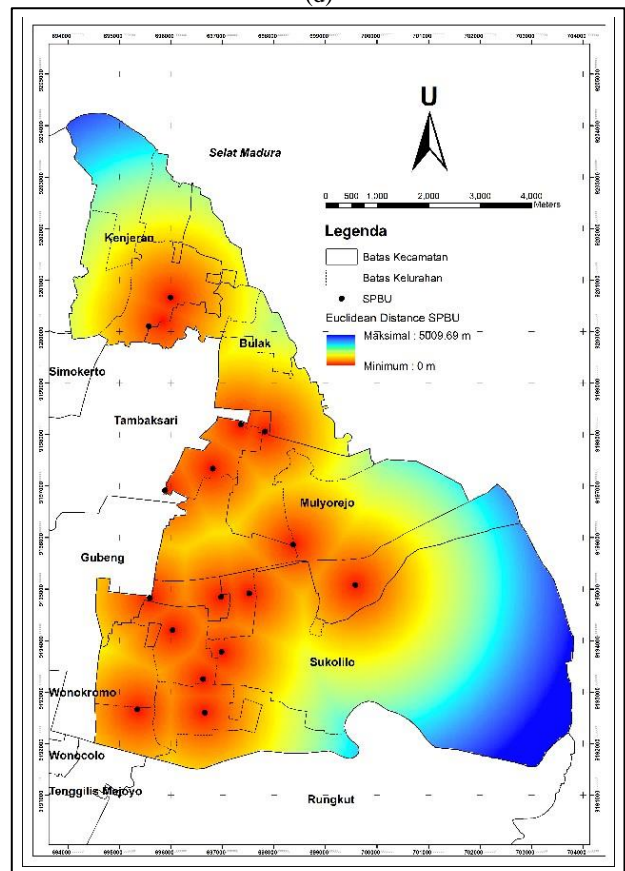
(c)



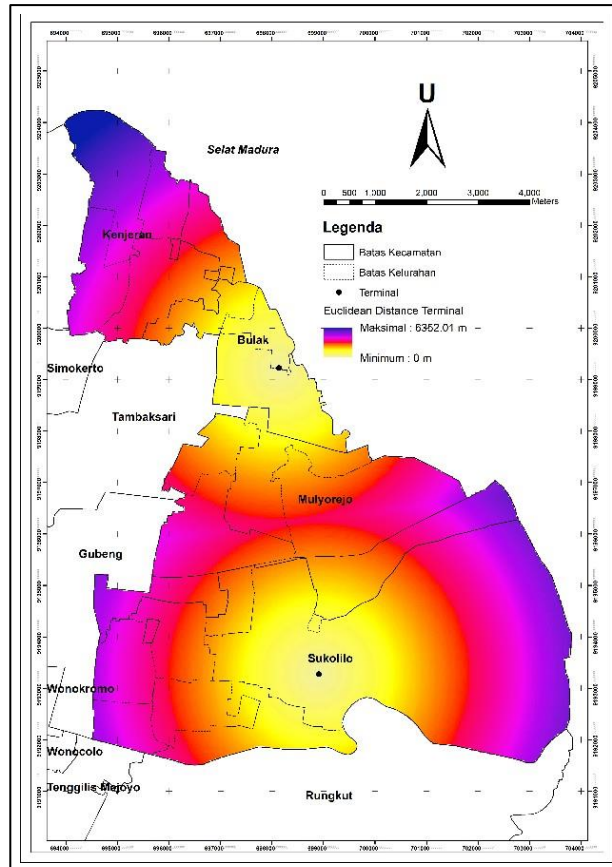
(d)



(e)



(f)



(g)

Gambar 2. Peta Parameter Jarak dalam Area Studi Kasus: (a) Fasilitas Kesehatan, (b) Objek Wisata, (c) Perbankan, (d) Pasar, (e) Mall, (f) SPBU, (g) Terminal

Tabel 1. Jumlah Parameter Fasilitas Sosial-Ekonomi

Parameter Fasilitas Sosial-Ekonomi	Jumlah
Fasilitas Kesehatan	49
Objek Wisata	34
Perbankan	193
Pasar	16
Mall	3
SPBU	16
Terminal	2

Untuk mendapatkan parameter jarak masing-masing sel dari obyek terdekat, maka dilakukan pengolahan euclidean distance dari fasilitas sosial-ekonomi tersebut yang diukur menggunakan spatial analyst tools. Berdasarkan pengolahan euclidean distance yang telah dilakukan pada setiap parameter didapatkan hasil parameter jarak, yaitu fasilitas kesehatan (rumah sakit, puskesmas, dan klinik) dengan jarak 0-4915,45m, objek wisata (tempat wisata dan taman kota) dengan jarak 0-3975,34m, perbankan (atm dan bank) dengan jarak 0- 3632,28m, pasar dengan jarak 0-4964,15m, mall dengan jarak 0-9166,83, SPBU dengan jarak 0-5009,69m, dan terminal dengan jarak 0-6352,01m yang divisualisasikan seperti pada Gambar 2.

2. Pemodelan Random Forest

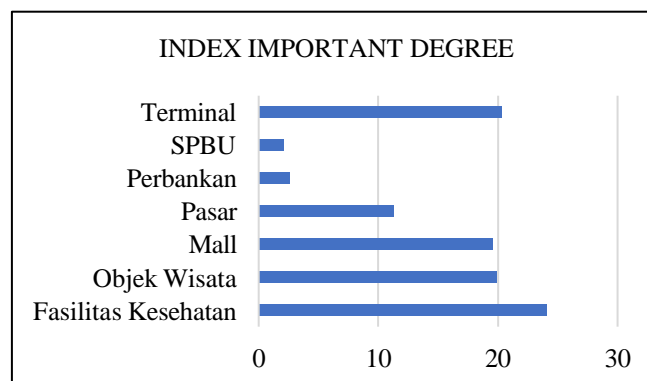
Setelah dilakukan proses pembuatan parameter jarak pada setiap fasilitas sosial-ekonomi menggunakan *euclidean distance tools*, dilakukan pemodelan *random forest* menggunakan 8712 data kejadian COVID-19 dan 8712 data non-kejadian COVID-19 dengan rincian seperti pada Tabel 2.

Kemudian, data tersebut dibagi menjadi dua, yaitu data *training* sebesar 70% dan data *testing* sebesar 30%.

Pada penelitian ini digunakan 12196 data *training* dan 5228 data *testing*. Pada tahap ini proses algoritma *random forest* digunakan untuk mendapatkan model dan menghitung pembobotan parameter yang akurat. Parameter ini memungkinkan pembuat keputusan untuk menentukan kontribusi indeks atau parameter terhadap penyebaran pandemi COVID-19 di Unit Pengembangan II Kertajaya dan III Kota Surabaya. Berdasarkan pengolahan *random forest* didapatkan tingkat pengaruh (*index importance degree*) dari setiap parameter terhadap penyebaran COVID-19, yaitu fasilitas kesehatan 24,135%, terminal 20,338%, objek wisata 19,916%, mall 19,574%, pasar 11,317%, perbankan 2,628%, dan SPBU 2,092% seperti pada Gambar 3.

Tabel 2. Data Kejadian Kasus COVID-19 pada Bulan Maret Tahun 2020 hingga Agustus Tahun 2021
(Sumber: Satuan Tugas COVID-19 Pemerintah Kota Surabaya)

Kecamatan	Kelurahan	Jumlah
Bulak	Kenjeran	192
	Bulak	458
	Kedung Cowek	49
Kenjeran	Sukolilo Baru	200
	Tanah Kali Kedinding	838
	Sidotopo Wetan	687
	Bulak Banteng	361
	Tambak Wedi	300
Mulyorejo	Mulyorejo	573
	Kejawen Putih Tambak	187
	Manyar Sabrangan	523
	Dukuh Sutorejo	447
	Kalisari	505
Sukolilo	Kalijudan	421
	Menur Pumpungan	429
	Semolowaru	554
	Nginden Jangkungan	346
	Klampus Ngasem	474
	Gebang Putih	150
	Keputih	535
	Medokan Semampir	483



Gambar 3. Tingkat Pengaruh (*Index Important Degree*) Setiap Parameter

Berdasarkan hasil *index importance degree* tersebut didapatkan bahwa parameter yang paling mempengaruhi penyebaran COVID-19 di Unit Pengembangan II Kertajaya dan III Tambak Wedi Kota Surabaya adalah fasilitas kesehatan dengan persentase sebesar 24,135% dan parameter yang paling tidak mempengaruhi adalah SPBU dengan persentase sebesar 2,092%.

3. Uji Akurasi Pemodelan Random forest

Pada penelitian ini dilakukan uji akurasi pengolahan *random forest* menggunakan *confusion matrix* dan kurva ROC. Uji akurasi merupakan tahapan yang bertujuan untuk mengevaluasi kinerja algoritma yang digunakan dalam penelitian selama proses pengolahan pada model *random forest* yang digunakan.

Confusion Matrix and Statistics			
		Reference	
Prediction	no	yes	
no	2440	108	
yes	174	2506	
Accuracy : 0.9461			
95% CI : (0.9396, 0.952)			
No Information Rate : 0.5			
P-value [Acc > NIR] : < 2.2e-16			
Kappa : 0.8921			
McNemar's Test P-value : 0.0001085			
Sensitivity : 0.9334			
Specificity : 0.9587			
Pos Pred Value : 0.9576			
Neg Pred Value : 0.9351			
Prevalence : 0.5000			
Detection Rate : 0.4667			
Detection Prevalence : 0.4874			
Balanced Accuracy : 0.9461			
'Positive' Class : no			

Gambar 4. Hasil Uji Akurasi Akhir Menggunakan Data Testing

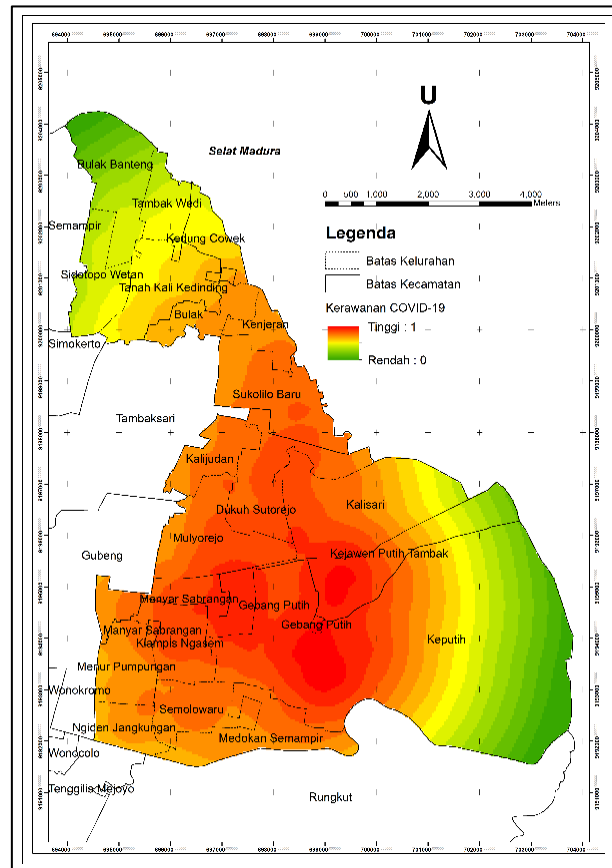
Confusion matrix adalah metode perhitungan untuk mengetahui dan melakukan analisis kualitas model klasifikasi yang terdiri dari *true positive* yaitu jumlah data positif yang terklasifikasi dengan benar, *false positive* yaitu jumlah data negatif yang terklasifikasi dengan benar, *false negative* yaitu jumlah data negatif namun terklasifikasi salah, *true negative* yaitu jumlah data positif namun terklasifikasi salah.

Evaluasi dengan *confusion matrix* menghasilkan nilai *accuracy*, *specificity*, dan *sensitivity* seperti pada Gambar 6. *Accuracy* adalah persentase ketepatan *record* data yang diklasifikasikan secara benar setelah dilakukan pengujian pada hasil klasifikasi (Han & Kamber, 2006). Sedangkan *specificity* adalah proporsi kasus negatif yang diprediksi negatif secara benar dan *sensitivity* adalah proporsi kasus positif yang diprediksi benar pada data yang sebenarnya (Powers, 2011). Pada penelitian ini didapatkan nilai *accuracy* sebesar 0,946 dan nilai kappa sebesar 0,892, dimana hasil tersebut tergolong pada kategori keakuratan yang baik. Kemudian, didapatkan juga nilai *sensitivity* sebesar 0,933 menunjukkan bahwa total 93,34% kejadian COVID-19 dideteksi dengan benar (*True Positive*) dan nilai *specificity* yang sebesar 0,9587 menunjukkan bahwa total 95,87% kejadian tidak COVID-19 dideteksi dengan benar (*True Negative*).

Kurva ROC menunjukkan akurasi klasifikasi secara visual berdasarkan hasil nilai *sensitivity* dan *specificity* yang didapatkan dan selanjutnya dihasilkan nilai AUC (*Area Under Curve*), dimana kurva AUC diurutkan ke dalam lima kelas, yaitu *poor* (0,5-0,6), *moderate* (0,6-0,7), *good* (0,7-0,8), *very good* (0,8-0,9), dan *excellent* (0,9-1,00) (Pourghasemi dkk, 2020). Nilai AUC yang mendekati 1 menunjukkan akurasi yang tinggi, sedangkan nilai yang mendekati 0,5 menunjukkan akurasi yang rendah. Pada penelitian kali ini didapatkan hasil kurva AUC sebesar 0,984 yang tergolong dalam kategori *excellent*.

4. Perhitungan Luas Area yang Terdampak Kebakaran

Setelah melakukan pengolahan model *random forest* dan mendapatkan tingkat pengaruh setiap parameter (*index important degree*), dilakukan pembuatan peta kerawanan COVID-19 dengan menggunakan perangkat lunak pengolahan data spasial. Klasifikasi peta didasarkan dari hasil tingkat pengaruh setiap parameter untuk dimasukkan dalam pembobotan pada *sum weighted tools*. Dari hasil pembobotan tersebut, didapatkan nilai dengan tingkat kerawanan rendah yang digambarkan dengan warna hijau, sedang yang digambarkan dengan warna kuning, hingga tinggi yang digambarkan dengan warna merah.



Gambar 5. Peta Kerawanan COVID-19

Berdasarkan hasil pemetaan tersebut, didapatkan daerah dengan tingkat kerawanan rendah pada bagian sebelah utara Kecamatan Kenjeran dan sebelah timur Kecamatan Sukolilo dan Kecamatan Mulyorejo, sedangkan tingkat kerawanan tinggi pada bagian tengah Kecamatan Mulyorejo dan bagian tengah Kecamatan Sukolilo.

Kesimpulan

Pada penelitian ini dapat disimpulkan bahwa pengolahan *random forest* tingkat pengaruh setiap fasilitas sosial-ekonomi terhadap penyebaran COVID-19 di Unit Pengembangan II Kertajaya dan Unit Pengembangan III Tambak Wedi Kota Surabaya menghasilkan nilai tingkat pengaruh fasilitas kesehatan 24,135%, terminal 20,338%, objek wisata 19,916%, mall 19,574%, pasar 11,317%, perbankan 2,628%, dan SPBU 2,092%. Berdasarkan pemodelan *random forest* yang telah dibuat, didapatkan hasil uji akurasi menggunakan *confusion matrix* dan kurva ROC sebagai evaluasi kinerja sistem. Pada penelitian ini, didapatkan nilai *accuracy* sebesar 0,946 dan nilai *kappa* sebesar 0,892. Hasil tersebut tergolong pada kategori keakuratan yang baik. Kemudian, didapatkan juga nilai *sensitivity* sebesar 0,933 menunjukkan bahwa total 93,34% dan nilai *specificity* yang sebesar 0,9587 menunjukkan bahwa total 95,87%, sehingga menghasilkan nilai AUC (*Area Under Curve*) sebesar 0,984 yang dapat disimpulkan bahwa akurasi ini tergolong dalam kategori *excellent*. Berdasarkan hasil *index important degree*, didapatkan daerah dengan tingkat kerawanan rendah terdapat pada bagian sebelah utara Kecamatan Kenjeran dan sebelah timur Kecamatan Sukolilo dan Kecamatan Mulyorejo, sedangkan tingkat kerawanan tinggi terdapat pada bagian tengah Kecamatan Mulyorejo, dan bagian tengah Kecamatan Sukolilo.

Ucapan Terimakasih

Penulis mengucapkan terimakasih kepada Satuan Tugas COVID-19 Pemerintah Kota Surabaya selaku penyedia data yang digunakan pada penelitian ini.

Daftar Pustaka

- Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., et al. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020, 395(10223):507–13.
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J. et al. (2020) A novel coronavirus from patients with pneumonia in China, 2019. *The New England Journal of Medicine*, 382, 727–733. doi: 10.1056/NEJMoa2001017.
- Brevik, E.C., Steffan, J.J., Rodrigo-Comino, J., Neubert, D., Burgess, L.C., Cerda, A. (2019). Connecting the public with soil to improve human health. *Eur J Soil Sci*, 70(4):898– 910
- Kumar, V., Parihar, R.D., Sharma, A., Bakshi, P., Singh, S.G.P., Bali, A.S., et al. (2019). Global evaluation of heavy metal content in surface water bodies: a meta-analysis using heavy metal pollution indices and multivariate statistical analyses. *Chemosphere* 2019, 236:124364. <http://dx.doi.org/10.1016/j.chemosphere.2019.124364>.
- Yadav, N., Singh, S., Goyal, S.K. (2020). Effect of seasonal variation on bacterial inhabitants and diversity in drinking water of an office building, Delhi. *Air Soil Water Res* 2020, 12(1). <http://dx.doi.org/10.1177/1178622119882335>.
- Brizuela, N.G.; García-Chan, N.; Gutiérrez Pulido, H.; Chowell, G. (2021). Understanding the role of urban design in disease spreading. *Proc. R. Soc. A* 2021, 477, 20200524.
- Wu, J.T., Leung, K., Leung, G.M. (2020). Articles nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*, 395(10225):689–97
- Breiman, L. (2011). *Random forests*. Netherlands: Kluwer Academic Publishers.
- Han, J., & Kamber, M. (2006). *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman. ISBN 13: 978-1-55860-901-3.
- Powers, D.M.W. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness& Correlation. *Journal of Machine Learning Technologies*, ISSN: 2229-3981 & ISSN: 2229-399X, Volume 2, Issue 1, 2011, pp37-63.
- Pourghasemi, Hamid, R. Bahram, H., Soheila, P., & Zakariya, F. (2020). *Spatial modelling, risk mapping, change detection, and outbreak trend analysis of coronavirus (COVID-19) in Iran (days between 19 February to 14 June 2020)*. *International Journal of Infectious Diseases*, 98:90-108.



This article is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).