



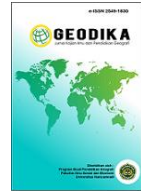
Website: <http://e-journal.hamzanwadi.ac.id/index.php/gdk>



**GEODIKA**  
Jurnal Kajian Ilmu dan Pendidikan Geografi

Terakreditasi S4 – SK No. 36/E/KPT/2019

Penerbit: Universitas Hamzanwadi



## ANALISIS DISTRIBUSI FASILITAS SEKOLAH MENENGAH PERTAMA MELALUI PEMODELAN SPASIAL STUDI KASUS DI KOTA MALANG

Kuncoro Adi Pradono<sup>1\*</sup>, Adi Wibowo<sup>2</sup>, Parluhutan Manurung<sup>3</sup>

<sup>1,2,3</sup>Departemen Geografi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Indonesia, Depok, Jawa Barat, Indonesia

\*Email Koresponden: [kuncoro.adi21@ui.ac.id](mailto:kuncoro.adi21@ui.ac.id)

Diterima: 30-07-2024, Revisi: 13-08-2024, Disetujui: 16-09-2024

©2024 Universitas Hamzanwadi

**Abstrak.** Dalam upaya membangun bangsa melalui pemerataan wajib belajar sembilan tahun merupakan bagian amanat UUD 1945 dan menjadi prioritas program pemerintah. Penelitian ini bertujuan untuk menganalisis distribusi fasilitas sekolah SMP di Kota Malang. Metode pemodelan spasial yang digunakan metode krigging dan regresi machine learning untuk prediksi perkembangan fasilitas sekolah SMP. Pemodelan spasial akan menilai sejauh mana pola spasial fasilitas sekolah terdistribusi dan analisis regresi yang akan melakukan prediksi fasilitas sekolah SMP dengan variabel peserta didik, guru, pegawai dan rombongan belajar. Hasil pemodelan spasial dengan teknik krigging menunjukkan distribusi variogram antar fasilitas sekolah berjarak 3-4 km dan merata. Adapun regresi dengan model terbaik secara berurutan REML, RF, OLS, GB, INLA dan PLS. Dengan model terbaik didapatkan akurasi 0,98 dan RMSE sebesar 0.79. Melalui hasil dari penelitian ini didapatkan gambaran bahwa sebaran distribusi fasilitas sekolah tingkat SMP di era zonasi saat ini masih terpusat di wilayah tengah Kota Malang sehingga terdapat peluang untuk pembangunan dan pengembangan fasilitas sekolah SMP di daerah pinggiran. Analisis pemodelan spasial dapat memberikan sudut pandang dan pertimbangan bagaimana fasilitas sekolah di perbaiki sesuai untuk pemerataan pendidikan.

**Kata kunci:** Fasilitas Sekolah, Pemodelan, Spasial, Regresi, Malang

**Abstract.** In an attempt to build the nation through equitable distribution of nine-year compulsory education is part of the mandate of the 1945 Constitution and is a priority program of the government. This research explores the distribution of junior high school facilities in Malang City, which is one of the cities of education. The purpose of this research is to consider the spatial distribution of junior high school education facilities in Malang City through spatial modeling with the kriging method and prediction of the development of junior high school facilities with machine learning regression. Spatial modeling will assess the extent to which the spatial pattern of school facilities is distributed, and regression analysis will predict junior high school facilities with variables of students, teachers, staff and study groups. The results of spatial modeling with kriging techniques show that the distribution of variograms between school facilities is within 3-4 km and evenly distributed. The regression with the best model is successively REML, RF, OLS, GB, INLA and PLS. With the best model obtained accuracy of 0.98 and RMSE of 0.79. Through the results of this study, it is found that the distribution of junior high school facilities in the current zoning era is still centered in the central area of Malang City, so there are opportunities for the construction and development of junior high school facilities in peripheral areas. Spatial modeling analysis can provide a point of view and consideration on how to improve school facilities according to educational equity.

**Keywords:** School Facility, Modeling, Spatial, Regression, Malang

### PENDAHULUAN

Upaya mencerdaskan kehidupan bangsa dengan pendidikan merupakan bagian amanat dari UUD 1945. Saat ini pemerintah memiliki program pendidikan dasar sembilan tahun sebagai salah satu kebijakan pemerintah dalam implemenasi UUD 1945. Hal tersebut dikuatkan melalui UU No 20 2003 yang disebutkan pada pasal 1 bagian ayat 6 dimana setiap bagian warga negara dengan ketentuan usia tujuh tahun hingga lima belas tahun diberikan kewajiban dalam mengeyam pendidikan tingkat dasar. Tujuan program pendidikan tingkat dasar sembilan tahun untuk memberikan kesempatan dan dorongan pada perkembangan potensi, hard skill dan soft skill peserta didik jenjang dasar sehingga

bertransformasi menjadi bekal insan unggul yang memiliki berwawasan kebangsaan, cinta tanah air, memiliki ketakwaan dan iman pada Tuhan, berbudi pekerti yang luhur, tubuh yang sehat, pikiran kreatif, semangat kemandirian sehingga dapat menjadi bagian warga negara yang berazaskan Pancasila (Bustari, 2016). Pemerataan fasilitas dan infrastruktur pendidikan dasar dilakukan pemerintah melalui program sekolah berdasarkan zonasi yang merepresentasikan kualitas pendidikan berkeadilan untuk warga negara sebagai bentuk pelayanan publik.

Salah satu tahapan tingkat dasar pendidikan di Indonesia yang dikenal dengan sekolah menengah pertama (SMP) atau juga dikenal sebagai Sekolah Lanjutan Tingkat pertama (SLTP) merupakan jenjang pendidikan dasar formal di Indonesia setelah lulus sekolah dasar. Umumnya SMP ditempuh dalam waktu 3 tahun namun ada pula yang 2 tahun seperti program akselerasi, mulai dari kelas 7 sampai kelas 9. SMP masuk dalam program wajib belajar 9 tahun. Fasilitas penunjang pendidikan dasar pada tingkat SMP perlu dipenuhi oleh pemerintah daerah selain tenaga pengajar dan kependidikan (Ramadhana, Sriyanto, & Prakoso, 2017). Salah satu aspirasi masyarakat yaitu mewujudkan pemerataan sebaran fasilitas pendidikan secara spasial yang cepat dan mudah sebagai wujud layanan pendidikan dimana merupakan fasilitas publik (Handayani & Wibowo, 2024). Terdapat permasalahan pemerataan fasilitas pendidikan sesuai dengan laju pertumbuhan penduduk di suatu wilayah. Adapun fasilitas tersebut seperti ruang kelas, beberapa ruang lab dan infrastruktur penunjang kegiatan belajar lain. Ada baiknya semua sekolah SMP memiliki fasilitas penunjang yang sama kuantitas dan kualitasnya di era zonasi sekolah saat ini sebagai wujud pemerataan kualitas pendidikan.

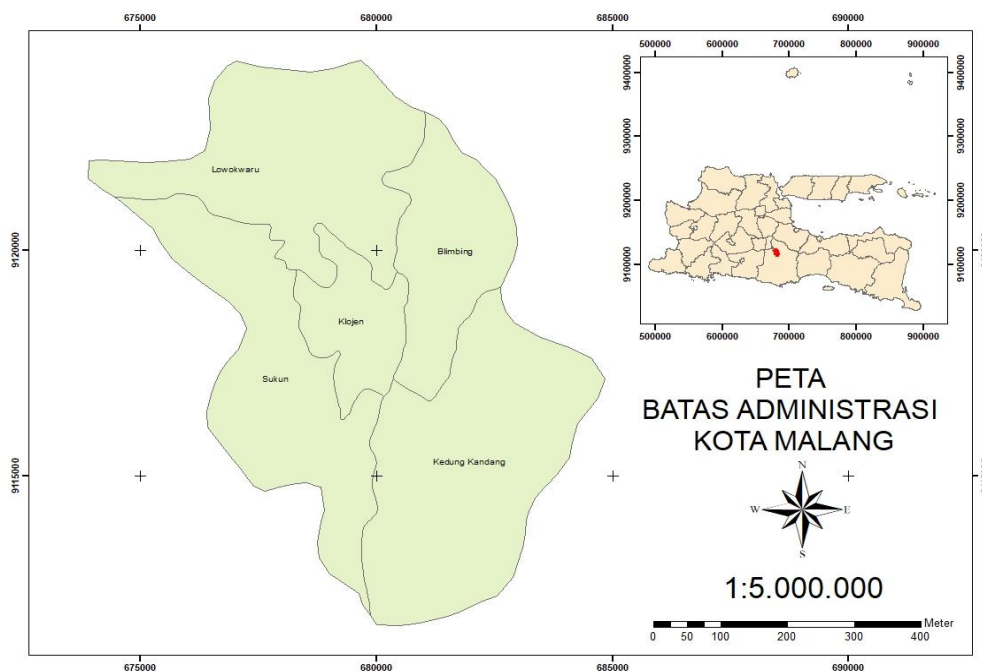
Kota Malang sebagai salah satu simpul sentra pendidikan di Indonesia karena banyak terdapat pusat pendidikan di berbagai jenjang (Sulaksono, 2017). Dalam jenjang pendidikan dasar dari data yang terdapat pada data pokok pendidikan (Dapodik) yang dimiliki Dirjen pendidikan usia dini, pendidikan tingkat dasar dan pendidikan tingkat menengah diperkirakan terdapat 272 sekolah SD dan 94 sekolah SMP. Dengan pola pemilihan sekolah berdasarkan zonasi, sekarang warga Kota Malang yang telah lulus SD diharapkan dapat memilih SMP terdekat. Tentu dalam memilih sekolah lanjutan salah satu kriteria yang menjadi acuan adalah fasilitas infrastruktur yang ada di SMP tersebut. Terkadang permasalahan muncul karena masyarakat masih memiliki kecenderungan memilih sekolah berdasarkan fasilitas unggulan seperti pada periode pendidikan sebelumnya sehingga melalui sistem pemilihan sekolah berdasarkan zonasi diharapkan dapat menginisiasi pemerataan distribusi kualitas dan fasilitas pendidikan. Laju pertumbuhan penduduk akan mendorong perkembangan kota, kebutuhan akan perkembangan sebaran sekolah juga akan meningkat. Masyarakat juga memiliki kecenderungan memilih sekolah dengan aksesibilitas yang mudah sehingga dalam pembangunan fasilitas sekolah tetap memperhatikan faktor fisik dan non fisik (Ainiyah & Wibowo, 2023). Dewan Perwakilan Rakyat Daerah kota Malang menginginkan pemerataan fasilitas dan sarana prasarana agar tidak terjadi permasalahan ketika penerimaan peserta didik baru (Politikmalang, 2023). Melalui pemanfaatan kemajuan teknologi spasial sistem informasi geografis (SIG) dan teknologi informasi, fasilitas sekolah SMP yang ada dapat diasosiasikan berbasis pemetaan secara geografis dan dibuatkan prediksi modelnya. Hal ini dapat membantu para pemangku kepentingan terkait sistem pendidikan SMP untuk melakukan perbaikan dan pemerataan fasilitas sekolah. Selain itu membantu warga Kota Malang mengorganisasikan informasi spasial distribusi fasilitas sekolah melalui SIG.

Pemanfaatan Sistem Informasi Geografis (SIG) saat ini mengalami perkembangan pesat dengan dukungan berbagai peralatan pemetaan, *mobile gadget* dan infrastruktur jaringan akibat adanya dukungan lompatan teknologi (Nikmawati, Bintoro, & Santoso, 2021). Selain itu era revolusi teknologi informasi 4.0 mendorong penggunaan *machine learning* dan berbagai model prediksi untuk membantu manusia agar lebih presisi dalam melakukan prediksi dan melakukan *decision support system* (DSS) (Syarovy, Nugroho & Sutiarmo, 2023). Beberapa algoritma *machine learning* digunakan untuk membantu melakukan prediksi distribusi spasial seperti krigging, random forest, *Ordinary Least Squares* (OLS), *Restricted Maximum Likelihood* (RMEL), *Partial Least Squares* (PLS), *Gradient Boosting* (GB) dan *Integrated Nested Laplace Approximation* (INLA) (Kuswanda, Damayanti & Ramadhani, 2024). Dengan ketersediaan peralatan, algoritma dan model dalam mendukung prediksi distribusi spasial tersebut dimungkinkan untuk dilakukan implementasi dalam melakukan analisis distribusi fasilitas sekolah untuk tingkat SMP.

Penelitian sebelumnya yang dilakukan oleh Sonhaji dkk mengenai permasalahan distribusi fasilitas sekolah menunjukkan pola persebarannya sesuai dengan pemukiman dan akses di Kab Gresik dengan metode linear (Sonhaji, Suprihardjo & Satiawan, 2010). Kemudian analisis distribusi fasilitas pelayanan sekolah tingkat menengah atas yang dilakukan Hamka dkk menunjukkan perbedaan tidak merata antara di daerah pedesaan dan perkotaan (Hamka, Missilia & Malik, 2022). Cakupan spasial wilayah sekolah dasar juga pernah diteliti oleh Harmes et al. dengan GIS melalui spasial buffering (Harmes & Febriasti Bahar, 2023). Penelitian ini akan melakukan pemodelan spasial melalui algoritma *machine learning* dan krigging yang memiliki keunggulan dalam prediksi dengan multivariabel untuk lokasi yang berdekatan yang diperlukan pada analisis prediksi distribusi fasilitas sekolah SMP di Kota Malang yang mana belum pernah dilakukan sebelumnya (Fathan, Sukmono & Firdaus, 2019). Pemodelan spasial ini dilakukan dengan data administrasi dan geocoding dari fasilitas pada setiap sekolah SMP yang terdapat di Kota Malang. Adapun tujuan dari penelitian ini untuk memberikan pertimbangan pemerataan sebaran spasial fasilitas pendidikan sekolah SMP di Kota Malang melalui pemodelan spasial dengan metode krigging dan model prediksi yang paling sesuai distribusi fasilitas sekolah SMP dengan regresi *machine learning* sehingga pemerataan dalam pendidikan bisa tercapai.

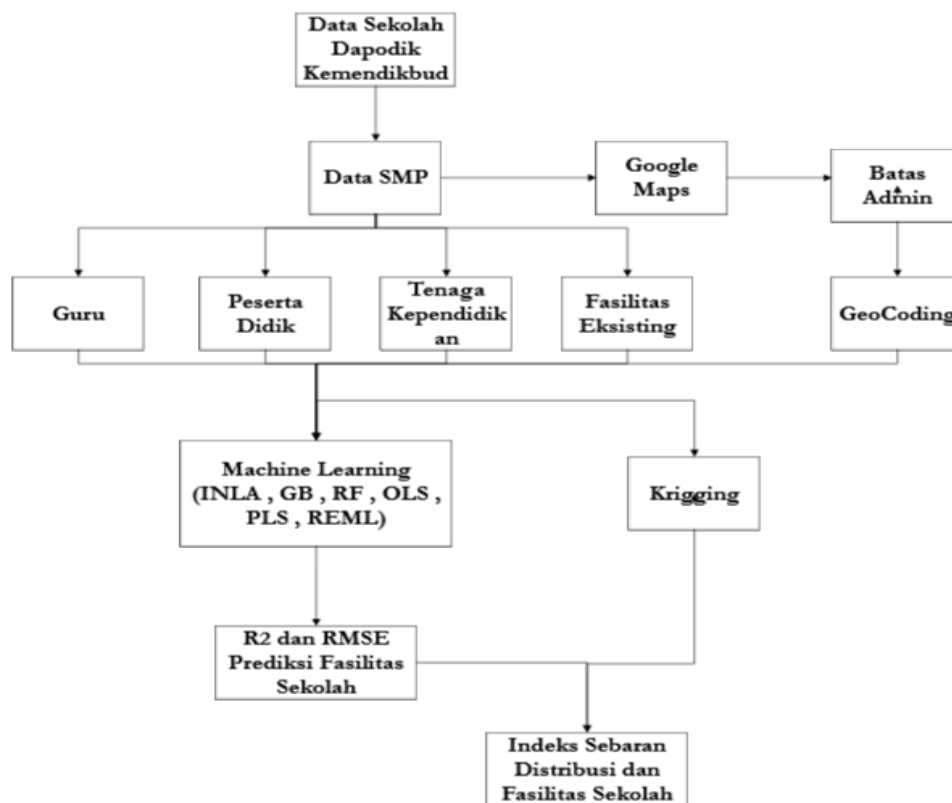
## METODE PENELITIAN

Penelitian disribusi fasilitas sekolah ini dilakukan untuk tingkat SMP yang berlokasi di Kota Malang. Secara spasial Kota Malang memiliki posisi diantara  $112,06^{\circ}$  -  $112,07^{\circ}$  Bujur Timur dan  $7,06^{\circ}$  -  $8,02^{\circ}$  Lintang Selatan. Selain itu dengan luasan sekitar 11.006 Ha terdapat pembagian administratif 5 Kecamatan dan 57 Kelurahan. Secara umum kondisi topografi datar dimana rentang kelerengn 0 - 15%. Adapun ketinggian yang diketahui berkisar antara 380 - 667 m di atas permukaan laut. Kota Malang dikenal sebagai kota pendidikan karena didukung banyak fasilitas diberbagai tingkat pendidikan di Kota seperti sekolah, Perguruan Tinggi Negeri (PTN) maupun Swasta (PTS) dan sekolah untuk pendidikan formal maupun non formal (Ramadanti, Mukhlis & Utomo, 2021).



**Gambar 1.** Peta Lokasi Penelitian  
(Sumber: Dokumen Peneliti, 2024)

Terdapat beberapa tahapan dalam penelitian ini, mulai dari tahap pengumpulan data, analisis, sampai penarikan kesimpulan penelitian. Adapun alur pikir penelitian dapat dilihat pada Gambar 2 yang disajikan dalam bentuk diagram alir.



**Gambar 2.** Alur Pikir Penelitian  
(Sumber: Dokumen Peneliti, 2024)

Penelitian ini menggunakan data spasial dan non spasial. Data administrasi Kota Malang yang diambil dari peta Rupa Bumi Indonesia tahun 2021 oleh Badan Informasi Geospasial. Data fasilitas, peserta didik, guru, pegawai dan rombongan belajar didapatkan dari Data Pokok Pendidikan (Dapodik) oleh Kementerian Pendidikan dan Kebudayaan. Berdasarkan alur pikir penelitian pada Gambar 2 dapat dijelaskan bahwa perangkat lunak yang digunakan yaitu *Opensource Software R Studio* untuk melakukan layout dan pemodelan spasial. Analisis spasial untuk memetakan kerapatan fasilitas sekolah SMP yang digunakan dengan pemodelan spasial teknik *Krigging*. Adapun untuk teknik regresi prediktif fasilitas sekolah SMP menggunakan Teknik *Ordinary Least Squares (OLS)*, *Restricted Maximum Likelihood (RMEL)*, *Partial Least Squares (PLS)*, *Gradient Boosting (GB)*, *Integrated Nested Laplace Approximation (INLA)* dan *Random Forest*.

**Tabel 1.** Tipe Iklim menurut Schmidt-Ferguson

No.	Data	Jenis Data	Keterangan
1	Batas Administrasi	Polygon Shapefile	Rupa Bumi Indonesia dari BIG
2	Data Lokasi	Point Shapefile	Geocoding
3	Peserta Didik	CSV	Dapodik Kemendikbud
4	Guru	CSV	Dapodik Kemendikbud
5	Pegawai	CSV	Dapodik Kemendikbud
6	Fasilitas	CSV	Dapodik Kemendikbud

Sumber: Rancangan Peneliti, 2024.

Berbagai metode telah diusulkan untuk menyesuaikan model regresi dengan estimasi dan prediksi data spasial yang berasal dari kerangka kerja yang berbeda: frequensial, Bayesian, dan berbasis Boosting (Efron & Hastie, 2016). *Kriging* adalah suatu metode atau model spasial dengan analisis data berbasis geostatistik untuk melakukan perkiraan besarnya nilai yang mewakili suatu titik dimana titik tersebut tidak tersampel berdasarkan titik yang telah tersampel di sekitarnya dengan menggunakan model struktural semivariogram. *Kriging* juga digunakan untuk menonjolkan metode khusus yang meminimalkan variansi dari hasil pendugaan (Fridayani, Kencana & Sukarsa, 2012). Teknik *Kriging* digunakan untuk melakukan estimasi distribusi fasilitas sekolah SMP yang ada di Kota Malang. Analisis spasial dari hasil kriging berupa interpolasi spasial yang mana digunakan untuk memodelkan distribusi fasilitas sekolah di wilayah Kota Malang untuk memprediksi jumlah fasilitas sekolah sehingga dapat memahami pola spasial dari variabel yang diamati. Prediktif dengan kriging  $\check{Z}(s)$  dimana kombinasi linear dari jumlah perkalian pasangan berurutan dari anggota himpunan dimana hal ini nilai lokasi fasilitas sekolah yang tidak tersampel. Prediktif kriging  $\check{Z}(s)$  yang diketahui ditulis melalui persamaan berikut:

$$\check{Z}(s) = \sum_a^n w_a Z(S_a)$$

Dimana:  $\check{Z}(s)$  merupakan nilai prediksi lokasi yang tidak pernah tersampel sebelumnya,  $w_i$  bagian dari koefisien bobot dari  $Z(S_a)$ ,  $n$  merupakan banyaknya sampel dan  $Z(S_a)$  merupakan nilai dimana lokasi telah tersampel.

Penelitian distribusi fasilitas sekolah SMP ini juga menggunakan pendekatan metode *Ordinary Least Squares (OLS)*. *OLS* adalah metode regresi linear yang digunakan untuk menemukan pola yang berkorelasi pada hubungan variabel dependen (terikat) dan satu atau lebih variabel independen (bebas). Untuk metode regresi dengan linier (Frequentist atau Bayesian), model polinomial orde pertama dan kedua dievaluasi untuk mengidentifikasi kecocokan terbaik dari struktur rata-rata model (Giannini Kurina, Hang, Macchiavelli & Balzarini, 2019). Metode ini memiliki keandalan dalam menganalisis faktor-faktor yang mempengaruhi distribusi fasilitas sekolah SMP di Kota Malang sesuai dengan ketersediaan data yang berhubungan dengan peserta didik, guru dan pegawai. Metode *OLS* dapat digunakan untuk menemukan pola korelasi yang paling sesuai antara variabel independen misalnya  $X$  dan variabel dependen misalnya  $Y$ , pada kasus ini variabel independen yaitu warga sekolah seperti peserta didik, guru dan pegawai, adapun variabel independen berupa fasilitas sekolah. *OLS* akan dimanfaatkan untuk mempertimbangkan kekuatan pola korelasi dan melakukan prediksi terhadap variabel dependen melalui variabel independen yang diamati (Walpole & R. H., 2012). Metode prediksi parameter yang berpengaruh pada regresi *OLS* dilakukan secara matematis melalui Metode Kuadrat Terkecil (MKT). Model analisis regresi linier berganda akan menggunakan persamaan berikut:

$$Y = \beta_0 + \sum_{i=1}^p \beta_i X_i$$

Dimana:  $Y$  merupakan hasil dari respon,  $X$  merupakan faktor independen (prediktor) yang mempengaruhi,  $\beta_0$  merupakan intercept dan  $\beta_i$  merupakan parameter peubah ke  $i$  sampai dengan  $p$ .

Dalam penelitian ini, metode *Restricted Maximum Likelihood (REML)* digunakan untuk menemukan pola korelasi antara jumlah fasilitas sekolah dengan berbagai macam variabel prediktor seperti jumlah peserta didik, guru, dan pegawai di wilayah Kota Malang. Metode *REML* dipilih karena kemampuannya dalam menangani efek acak (*random effects*) yang mungkin terdapat dalam data (O'Neill, 2013). Diharapkan estimasi yang dihasilkan jauh lebih baik dibandingkan dengan model regresi linier yang biasa digunakan. Untuk metode *Restricted Maximum Likelihood (REML)* dengan data yang meliputi jumlah fasilitas sekolah, jumlah peserta didik, guru, dan pegawai, serta koordinat geografis lokasi fasilitas sekolah, formula *REML* dapat ditulis sebagai berikut:

$$Y = X\beta + Zu + \varepsilon$$

Dimana:  $\beta$  vektor parameter *fixed effects*,  $u$  adalah vektor efek acak dengan  $u \sim N(0, \sigma^2 u)$ ,  $\varepsilon$  adalah vektor error dengan  $\varepsilon \sim N(0, \sigma^2 \varepsilon)$ . Dalam metode *REML*, estimasi parameter dilakukan dengan memaksimalkan kemungkinan terbatas (*restricted likelihood*) yang diperoleh dengan mentransformasi vektor respons  $y$  menjadi  $y^* = A'y$ , di mana  $A$  adalah matriks proyeksi yang memproyeksikan  $y$  ke ruang nol dari  $X$ , yaitu  $A = I - X(X'X)^{-1}X'$ .

Metode *Partial Least Squares (PLS)* digunakan sebagai teknik regresi untuk memodelkan hubungan antara jumlah fasilitas sekolah dengan variabel-variabel prediktor seperti jumlah peserta didik, guru, dan pegawai di Kota Malang. Metode *PLS* dipilih karena keandalannya dalam menangani masalah multikolinearitas, yaitu pola hubungan yang tinggi antara berbagai macam variabel prediktor (Supriyadi, Mariani & Juli, 2017). Dalam *PLS*, variabel respons  $Y$  dan variabel prediktor  $X$  diproyeksikan ke ruang berdimensi rendah yang dibentuk oleh komponen laten atau faktor, yang disebut variabel laten (*latent variables*). Misalkan  $T$  adalah matriks variabel laten untuk  $X$  berukuran  $n \times k$ , dan  $U$  adalah vektor variabel laten untuk  $y$  berukuran  $n \times 1$ , dengan  $k$  adalah jumlah komponen laten yang digunakan dalam model. Hubungan antara  $X$ ,  $T$ ,  $Y$ , dan  $U$  dalam model *PLS* dapat ditulis sebagai:

$$X = TG' + E$$

$$Y = Ub' + f$$

Dimana: variabel  $G$  dan  $b$  bagian dari vektor bobot untuk  $X$  dan  $Y$ , sedangkan  $E$  dan  $f$  adalah error atau residu. Vektor bobot  $P$  dan  $c$  diperoleh dengan memaksimalkan kovarian antara  $T$  dan  $U$ , atau secara ekuivalen, memaksimalkan korelasi antara  $y$  dan  $Xw$ , dengan  $w$  adalah vektor bobot untuk  $X$ .

Metode *Gradient Boosting (GB)* merupakan pendekatan *machine learning* untuk memodelkan hubungan antara jumlah fasilitas sekolah dengan variabel-variabel prediktor seperti jumlah peserta didik, guru, dan pegawai di Kota Malang. Metode *GB* memiliki kemampuan dalam menangani hubungan non-linier dan interaksi antara variabel-variabel prediktor (Laili, Purwarianti, 2022). Dalam *GB*, model prediksi akhir  $F(x)$  dibangun dengan menggabungkan banyak model pohon keputusan sederhana  $h(x; a_m)$  yang dilatih secara bertahap (iteratif) untuk meminimalkan kesalahan prediksi residual dari model sebelumnya. Formulasi *GB* dapat ditulis sebagai:

$$F(x) = F_0(x) + \sum 1^{M \rho_m} h(x; a_m)$$

Dimana:  $F(x)$  adalah model prediksi akhir,  $F_0(x)$  adalah nilai awal prediksi (biasanya konstanta),  $M$  adalah jumlah iterasi boosting,  $M \rho_m$  adalah *shrinkage factor* pada iterasi ke- $m$  (biasanya bernilai kecil, misalnya 0,1),  $h(x; a_m)$  adalah model pohon keputusan sederhana dengan parameter  $a_m$  pada iterasi ke- $m$ .

*Integrated Nested Laplace Approximation (INLA)* memiliki kehandalan untuk menunjukkan pola antara jumlah fasilitas sekolah dengan variabel-variabel prediktor seperti jumlah peserta didik, guru, dan pegawai di wilayah Kota Malang. Metode *INLA* memiliki kemampuan dalam menangani efek spasial dan efek acak lainnya yang mungkin terdapat dalam data, serta melakukan inferensi Bayesian secara efisien. Model *INLA* untuk data ini dapat diformulasikan sebagai berikut:

$$y_a \sim \text{Fungsi Poisson}(\mu_a)$$

$$\log(\mu_a) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + f_{\text{spat}}(s_i)$$

Dimana:  $y_a$  mengikuti distribusi fungsi Poisson dengan mean  $\mu_a$ ,  $\alpha$  merupakan intercept,  $\beta_1 x_{i1}$ ,  $\beta_2 x_{i2}$ ,  $\beta_3 x_{i3}$  merupakan koefisien regresi untuk variabel prediktor jumlah peserta didik, guru, dan pegawai,  $f_{\text{spat}}(s_i)$  adalah efek spasial pada lokasi  $s_i$ , yang dimodelkan dengan menggunakan struktur kovarians spasial. Efek spasial  $f_{\text{spat}}(s_i)$  biasanya dimodelkan dengan menggunakan model *Gaussian Markov Random Field (GMRF)*, yang mengasumsikan bahwa efek spasial pada suatu lokasi hanya bergantung pada efek spasial di lokasi-lokasi tetangga terdekat. Dalam *INLA*, distribusi posterior untuk parameter alfa, betha dan efek spasial diperkirakan dengan menggunakan metode pendekatan *Laplace Tersing (Nested Laplace Approximation)*. Pendekatan ini melibatkan penggunaan integrasi numerik dan teknik penghitungan Laplace untuk memperkirakan distribusi posterior secara efisien. Mengenai

kesalahan prediksi global, perhitungan regresi Bayesian untuk spasialitas menghasilkan kecocokan yang terbaik. Metode ini cocok untuk memahami proses retensi dan pemetaan. R-INLA menangani model Gaussian laten di mana efek tetap, terstruktur dan tidak terstruktur, efek acak Gaussian digabungkan secara linear (Bakka et al., 2018).

Dalam kasus data yang melibatkan koordinat spasial, informasi lokasi dapat digunakan sebagai salah satu fitur dalam proses pembangunan pohon keputusan. *Algoritma Random Forest* bekerja dengan metode pembelajaran ensemble untuk klasifikasi, regresi, dan melakukan prediksi lain melalui pembangunan banyak pohon keputusan pada waktu pelatihan (Belgiu & Drăguț, 2016). Hal ini memungkinkan *Random Forest* untuk menangkap hubungan antara peserta didik, guru, pegawai dan jumlah fasilitas sekolah. Pada *Random Forest*, setiap pohon keputusan  $T_b(x)$  dibangun dengan menggunakan subset data dan fitur yang dipilih secara acak. Formulasi untuk sebuah pohon keputusan tunggal  $T_b(x)$  dapat ditulis sebagai:

$$T_b(x) = \sum 1^{M_b} c_m I(x \in R_m)$$

Dimana:  $M_b$  adalah jumlah simpul terminal (*leaf nodes*) pada pohon ke- $b$ ,  $R_m$  adalah wilayah (region) yang sesuai dengan simpul terminal ke- $m$ ,  $c_m$  adalah nilai prediksi (jumlah fasilitas sekolah) untuk simpul terminal ke- $m$ ,  $I(x \in R_m)$  merupakan fungsi indikator dengan nilai 1 apabila  $x$  berada di wilayah  $R_m$ , dan 0 jika tidak, Setiap pohon keputusan  $T_b(x)$  dibangun dengan menggunakan subset data dan fitur yang dipilih secara acak. Subset data dipilih dengan teknik bootstrap, sedangkan subset fitur dipilih secara acak pada setiap simpul pemisahan (*splitting node*). Formulasi untuk prediksi akhir dari *Random Forest*  $f(x)$  adalah:

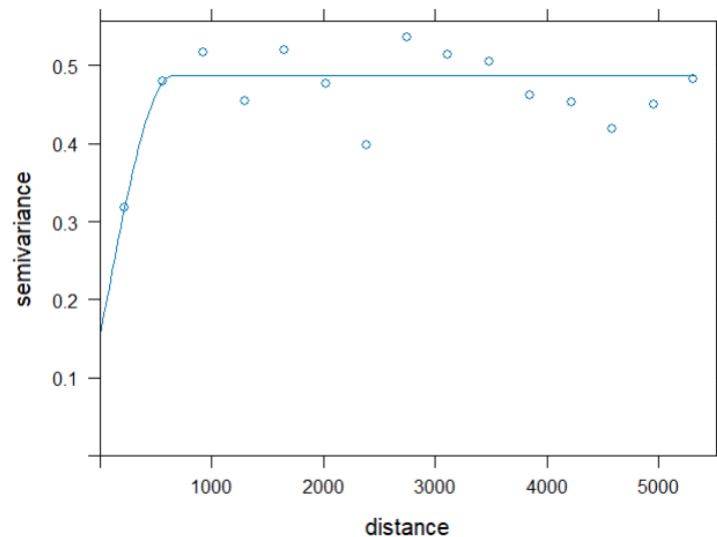
$$f(x) = \frac{\sum 1^{B} T_b(x)}{B}$$

Dimana:  $B$  adalah jumlah pohon dalam hutan (*forest*),  $T_b(x)$  adalah prediksi dari pohon keputusan ke- $b$ . Prediksi akhir  $f(x)$  diperoleh dengan mengambil rata-rata prediksi dari semua pohon keputusan dalam hutan.

## TEMUAN DAN PEMBAHASAN

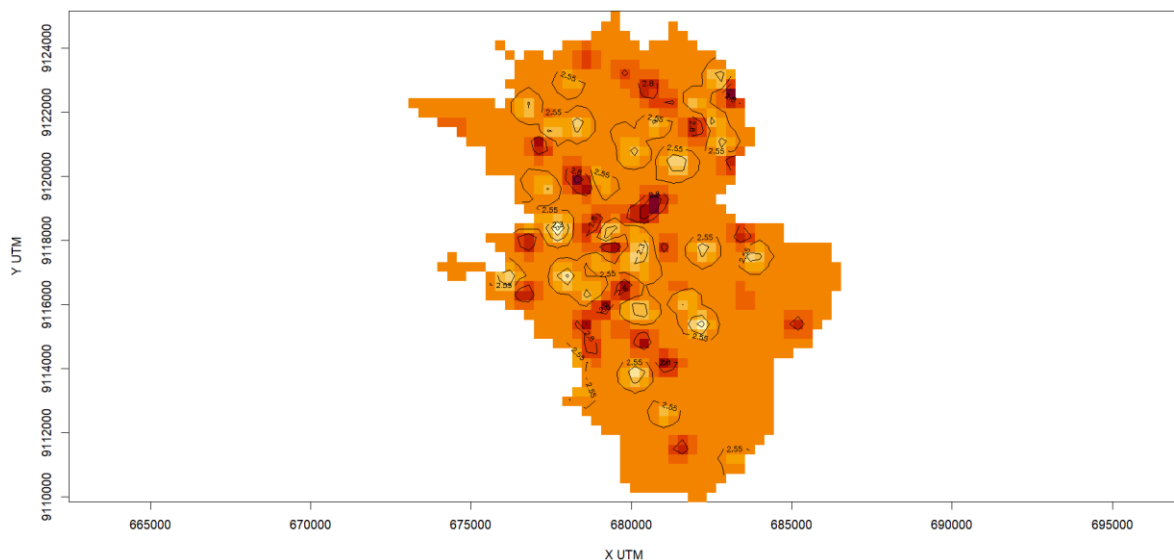
Berdasarkan hasil analisis spasial dengan teknik krigging dengan R studio yang telah dilakukan pada data fasilitas dengan sekolah yang tergeocoding model Gaussian (Gau), Spherical (Sph), Matern (Mat), dan Exponential (Exp). Model yang paling sesuai dengan data spasial distribusi fasilitas sekolah SMP Spherical (Sph) dengan adanya Nugget (Nug), dimana menunjukkan adanya variasi pada jarak yang dekat serta autokorelasi spasial yang kuat. Nilai psill yang relatif rendah 0.1541 hingga 0.3335 menunjukkan variasi yang tidak terlalu besar dalam distribusi fasilitas sekolah berarti jumlah fasilitas antar sekolah SMP relatif sama. Hal ini didukung dengan nilai range yang cukup besar 648.8 menandakan autokorelasi spasial yang meluas berarti fasilitas sekolah SMP berdekatan cenderung memiliki karakteristik yang mirip. Anisotropi bernilai 0 atau 1 menandakan variasi spasial seragam disemua area.

Adapun hasil semivariogram menggambarkan jarak pada data spasial. Semivariance cenderung rendah menandakan autokorelasi spasial yang kuat. Jarak variance antara 3.000 sampai 4.000 unit menunjukkan jarak antar fasilitas paling jauh 4 kilometer. Efek nugget yang kecil juga mendukung bahwa distribusi spasial fasilitas sekolah berada pada jarak yang dekat. Untuk lebih detail semigariogram ditunjukkan pada Gambar 3.



**Gambar 3.** Semivariogram krigging untuk fasilitas sekolah SMP di Kota Malang  
(Sumber: Hasil analisis data penelitian, 2024)

Dari hasil penelitian, didapatkan distribusi fasilitas yang ada di Kota Malang dengan indeks tinggi sampai dengan 2.8 yang terpusat di wilayah tengah Kota Malang. Walaupun tidak semua sekolah SMP memiliki fasilitas yang sama namun distribusi spasial hampir merata sampai ke wilayah pinggir Kota Malang, yang mana sekarang banyak permukiman baru sehingga peserta didik di wilayah tersebut juga mendapat fasilitas yang setara. Untuk fasilitas sekolah SMP yang tertinggi memang hampir terkumpul di pusat kota hal ini mungkin sekolah tersebut sudah lama berdiri sehingga banyak mengalami renovasi dan pengembangan fasilitas. Adapun peta distribusi spasial sebaran fasilitas sekolah SMP dari model Krigging dapat dilihat pada Gambar 4.

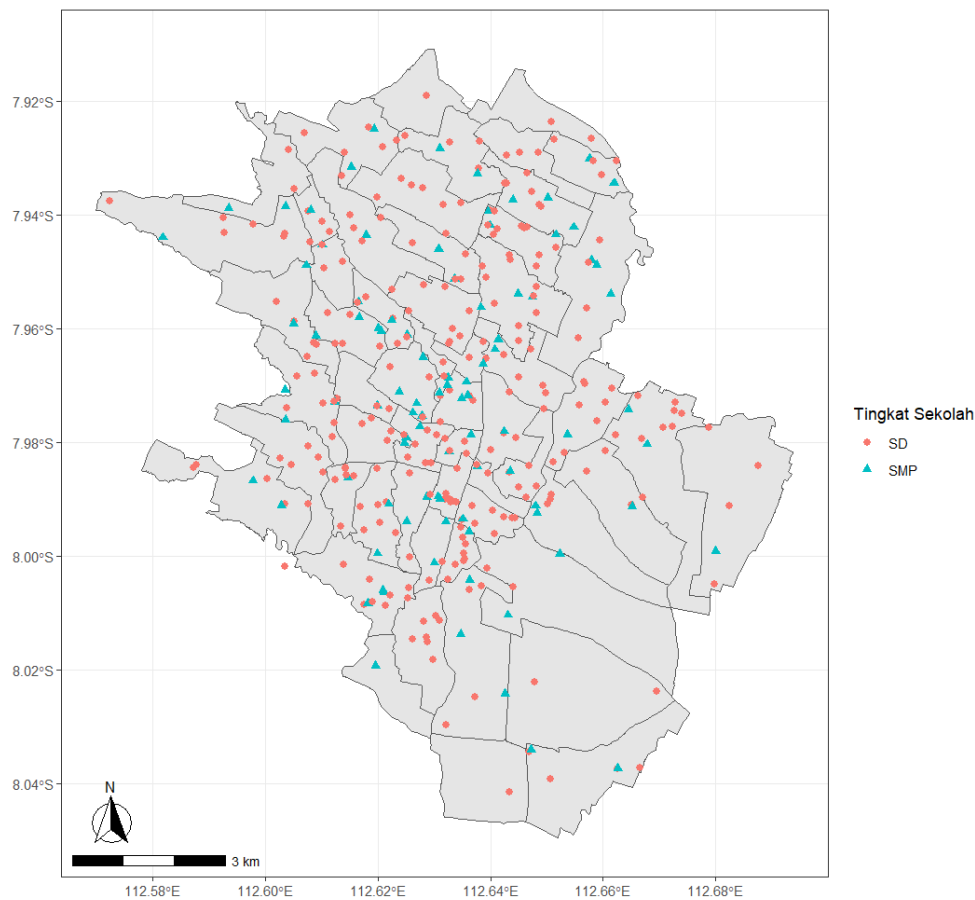


**Gambar 4.** Distribusi spasial sebaran fasilitas sekolah dengan metode Krigging  
(Sumber: Hasil analisis data penelitian, 2024)

Regresi dengan metode *OLS*, *REML*, *PLS*, *GB*, *INLA* dan *Random Forest* telah dilakukan untuk data fasilitas sekolah SMP terhadap data-data dari peserta didik, guru, pegawai dan rombongan belajar. Semua regresi ini dilakukan dengan membaca big data pendidikan melalui R studio. Pustaka untuk melakukan regresi tersedia secara *opensource*. Hasil dari perhitungan untuk setiap model pada  $R^2$  juga *Root Mean Square Error (RMSE)* ditampilkan pada Tabel 1. Metode *OLS* merupakan yang paling sederhana dan memiliki koefisien akurasi yang cukup baik pada 93%. Adapun metode dengan akurasi terbaik



untuk studi kasus ini dengan *REML* pada 97%. Variansi dari variabel dependen yang banyak seperti pada studi kasus ini dapat jelaskan hingga 90% dengan menggunakan metode *Random Forest*. Metode *PLS* memiliki akurasi paling rendah dengan 75% namun dapat menjelaskan variabel yang paling berpengaruh terhadap jumlah fasilitas sekolah SMP yaitu pegawai dan rombongan belajar. Metode *GB* memiliki akurasi 92% dengan keunggulan dapat menjelaskan korelasi antar variabel. Regresi dengan metode *INLA* menunjukkan akurasi 0.86% dan *RMSE* paling tinggi. Hasil mean hyperparameter dari metode *INLA* 0.128 menunjukkan presisi yang dapat dijelaskan dari tiap unit variasi variabel yang digunakan seperti peserta didik, guru, pegawai dan rombongan belajar. Selain itu hasil sebaran sekolah dalam pemodelan yang dijalankan dengan *R Studio* menunjukkan sebaran distribusi SMP masih terpusat dan berkembang ke daerah perbatasan di wilayah kota Malang. Titik sekolah dasar yang tersebar berdekatan dengan beberapa lokasi sekolah SMP sehingga siswa yang melanjutkan ke jenjang pendidikan SMP masih dapat menjangkau.



**Gambar 5.** Distribusi sekolah SD dan SMP dari hasil pemodelan *R Studio* dari data dapodik tahun 2024 (Sumber: Hasil analisis data penelitian, 2024)

Hasil pemodelan spasial secara umum menunjukkan persebaran fasilitas sekolah SMP yang merata. Dari hasil pemodelan, diketahui bahwa wilayah yang perlu peningkatan melalui pembangunan dan renovasi berada pada bagian wilayah pinggiran kota. Adapun dengan semua metode regresi hasil prediksi untuk distribusi fasilitas sekolah SMP sudah cukup baik sehingga dapat digunakan sebagai bahan pertimbangan berapa fasilitas yang akan dibangun untuk mencapai pemerataan pendidikan dari segi fasilitas di Kota Malang. Keterpusatan fasilitas sekolah SMP yang terlihat mungkin disebabkan kurikulum lama dalam pengembangan fasilitas sekolah berdasarkan banyaknya calon siswa atau siswa yang memilih bersekolah di sekolah unggulan yang mayoritas banyak terletak di pusat Kota Malang.

**Tabel 1.** Hasil regresi metode yang digunakan

No	Metode	R Square	RMSE
1	OLS	0.93732	2.80396
2	REML	0.98493	0.79690
3	PLS	0.75509	27.71291
4	GB	0.92221	7.57928
5	INLA	0.86303	15.64123
6	RF	0.95001	1.67073

Sumber: Hasil Analisis data penelitian, 2024.

## SIMPULAN

Pemodelan spasial distribusi fasilitas sekolah SMP di Kota Malang telah berhasil dilakukan dengan big data dapodik Kemendikbud dan geocoding dengan bantuan R Studio. Analisis pemodelan spasial menunjukkan pemerataan fasilitas sekolah SMP masih terpusat di wilayah pusat kota dan mulai berkembang secara menyebar melalui hasil pemodelan spasial dengan metode krigging yang ditunjukkan nilai 2.8. Adapun prediksi fasilitas sekolah dari metode regresi *machine learning* yang diterapkan umumnya menunjukkan akurasi yang baik. R square terbaik untuk prediksi dengan nilai 98% menggunakan metode *REML*. Hasil prediksi dengan *machine learning* yang secara umum di atas 90% dapat digunakan pemangku kepentingan untuk studi kelayakan pemerataan dalam pengembangan distribusi fasilitas sekolah khususnya tingkat SMP. Melalui hasil pemodelan spasial yang menunjukkan distribusi sekolah SMP pengembangan infrastruktur dapat dilakukan di daerah yang berada di pinggiran kota Malang. Melalui penelitian ini diharapkan pembangunan manusia melalui pendidikan dapat ditingkatkan dengan menggunakan analisis pemodelan spasial maupun saran untuk melakukan perbaikan dan peningkatan mutu pendidikan dari segi distribusi sekolah di Kota Malang untuk pemerataan.

## UCAPAN TERIMAKASIH

Ucapan terima kasih penulis sampaikan kepada Departemen Geografi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Indonesia, yang sudah memberikan kontribusi dalam terwujudnya penelitian ini.

## DAFTAR PUSTAKA

- Ainiyah, R., & Wibowo, A. (2023). Analisis Spasial Penentuan Lokasi Sekolah Baru SMA/SMK Sederajat (Studi Kasus: Kota Cilegon). *Geodika: Jurnal Kajian Ilmu dan Pendidikan Geografi*, 7(2), 175-185.
- Bakka, H., Rue, H., Fuglstad, G. A., Riebler, A., Bolin, D., Illian, J., ... & Lindgren, F. (2018). Spatial modeling with R-INLA: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(6), e1443.
- Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24–31.
- Bustari, M. (2016). Optimalisasi rehabilitasi ruang kelas dalam mendukung penyelenggaraan wajib belajar 9 tahun. *Jurnal Manajemen Pendidikan UNY*, 12(2), 113156.
- Efron, B., & Hastie, T. (2016). *Computer Age Statistical Inference*. Cambridge University Press.
- Fathan, M., Sukmono, A., & Firdaus, H. S. (2019). Analisis Kesesuaian Lahan Komoditas Kehutanan dan Pertanian di Wilayah Kabupaten Semarang dengan Metode Matching. *Jurnal Geodesi UNDIP*, 8(3), 8-16.
- Giannini Kurina, F., Hang, S., Macchiavelli, R., & Balzarini, M. (2019). Spatial Predictive Modelling Essential to Assess the Environmental Impacts of Herbicides. *Geoderma*, 354, 113874.

- Hamka, H., Misilia, M., & Malik, R. (2022). Analisis Distribusi Fasilitas Pelayanan Pendidikan Sekolah Menengah Atas (SMA). *Madani Jurnal Politik dan Sosial Masyarakat*, 14(03), 484-492.
- Handayani, A., & Wibowo, A. (2024). Analisis Spasial Sebaran Sekolah di Kecamatan Warungkondang Kabupaten Cianjur Provinsi Jawa Barat. *Jurnal Spatial Wahana Komunikasi dan Informasi Geografi*, 24(1), 66-71.
- Kuswanda, G. F., Damayanti, J., & Ramadhani, M. A. (2024). Pemodelan Semivariogram pada Data Potensi Calon Mahasiswa Baru Fakultas Teknik Sipil dan Perencanaan Universitas Trisakti. *Jurnal Penelitian dan Karya Ilmiah Lembaga Penelitian Universitas Trisakti*, 133-146.
- Yuhana, U. L., Purwarianti, A., & Imamah, I. (2022). Tuning Hyperparameter pada Gradient Boosting untuk Klasifikasi Soal Cerita Otomatis. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 8(1), 134-139.
- Nikmawati, N., Bintoro, H. S., & Santoso, S. (2021). Dampak Penggunaan Gadget terhadap Hasil Belajar dan Minat Belajar Siswa Sekolah Dasar. *Jurnal Edutech Undiksha*, 9(2), 254.
- O'Neill, M. (2013). *Matematika REML A workshop conducted*. Malang. Retrieved from [www.stats.net.au](http://www.stats.net.au)
- Politikmalang (2023). DPRD Kota Malang Dorong Pemerataan Fasilitas Pendidikan. Diakses pada 7 April 2024, dari <https://politikamalang.com/01/12/2023/dprd-kota-malang-dorong-pemerataan-fasilitas-pendidikan/yforce.co.uk/about-us/news/ideas-for-taking-geography-outside>
- Ramadanti, E., Mukhlis, I., & Hadi Utomo, S. (2021). Dampak pandemi COVID-19 terhadap Pendidikan Tinggi di Kota Malang. *Jurnal Ekonomi, Bisnis dan Pendidikan*, 1(3), 209-218.
- Ramadhana, A. N., Sriyanto, B., & Prakoso, E. (2017). Analisis Ketersediaan dan Keterjangkauan Fasilitas Pendidikan Jenjang Sekolah Dasar (SD) Dan Sekolah Menengah Pertama (SMP) Di Kota Metro.
- Kuswanda, G. F., Damayanti, J., & Ramadhani, M. A. (2024). Pemodelan Semivariogram Pada Data Potensi Calon Mahasiswa Baru Fakultas Teknik Sipil dan Perencanaan Universitas Trisakti. *Jurnal Penelitian dan Karya Ilmiah Lembaga Penelitian Universitas Trisakti*, 133-146.
- Sonhaji, A. N., Suprihardjo, R., & Satiawan, P. R. (2017). Konsep Distribusi Fasilitas Pendidikan Berdasarkan Wilayah dan Pola Persebaran Permukiman di Kabupaten Gresik. *Jurnal Penataan Ruang*, 5(1).
- Sulaksono, A. G. (2017). Pemanfaatan Aplikasi Sistem Informasi Geografis (SIG) untuk Pemetaan SMK Kota Malang. *Journal of Information Technology and Computer Science*, 2(2), 224611.
- Suma Fridayani, N. M., Nila Kencana, I. P. E., & Sukarsa, K. G. (2012). Perbandingan Interpolasi Spasial dengan Metode Ordinary dan Robust Kriging pada Data Spasial Berpencilan (Studi Kasus: Curah Hujan di Kabupaten Karangasem). *E-Jurnal Matematika*.
- Supriyadi, E., Mariani, S., & Juli, d. (2017). Perbandingan metode *partial least square (PLS)* dan Principal Component Regression (PCR) untuk Mengatasi Multikolinearitas pada Model regresi linear berganda. Retrieved from <http://journal.unnes.ac.id/sju/index.php/ujm>
- Syarovy, M., Nugroho, A. P., & Sutiarmo, L. (2023). Pemanfaatan Model Neural Network Dalam Generasi Baru Pertanian Presisi Di Perkebunan Kelapa Sawit. *WARTA Pusat Penelitian Kelapa Sawit*, 28(1), 39-54.
- Walpole, R., E. M., & R. H., M. (2012). *Probability & Statistics for Engineers & Scientist. 9th Edition* (9th ed.). Boston: Pearson Education, Inc.