

MODEL PENILAIAN TANAH MASSAL BERBASIS BIDANG TANAH MENGGUNAKAN ALGORITMA *RANDOM FOREST* DI KOTA SURAKARTA

PARCEL-BASED MASS VALUATION USING RANDOM FOREST IN SURAKARTA CITY

Catur Yulianto¹

¹Direktorat Penilaian Tanah dan Ekonomi Pertanahan, Kementerian ATR/BPN, Jakarta, Indonesia

Koresponden email: caturyulianto09@gmail.com

ABSTRAK

Penilaian tanah massal berbasis bidang tanah dapat dilakukan dengan menggunakan model berdasarkan variabel-variabel bebas pembentuk nilai tanah. Salah satu model yang disarankan untuk model nilai tanah yaitu algoritma *random forest*. Agar *random forest* memiliki kemampuan prediksi yang baik maka perlu dilakukan pengaturan *hyperparameter* antara lain *mtry*, jumlah pohon yang dibangun (*num.trees*), dan jumlah minimal *node* (*min.node.size*). Penelitian ini bertujuan untuk melihat kinerja model penilaian tanah menggunakan *random forest* dengan analisis pengaturan *hyperparameter* agar mendapatkan hasil akurasi terbaik. Hasil penelitian menunjukkan model nilai tanah memiliki nilai R^2 sebesar 82,96% dan nilai *mean absolute percent error* (MAPE) sebesar 26,83 sehingga model memiliki kemampuan prediksi yang layak. Berdasarkan standar rasio yang dikeluarkan oleh IAAO, uji kualitas model penilaian tanah menghasilkan nilai *coefficient of variation* (COV) sebesar 22,62 dan nilai *coefficient of dispersion* (COD) sebesar 25,29 serta nilai *price related differential* (PRD) sebesar 1,12. Nilai COV, COD, dan PRD masih di luar batas toleransi yang ditetapkan.

Kata kunci : model, penilaian tanah, *random forest*, *hyperparameter*

ABSTRACT

Parcel-based mass valuation can be carried out using a model based on the land value independent variables. One model that is recommended for modeling land value is the random forest algorithm. In order for random forest to have good predictive capabilities, it is necessary to tune hyperparameters including mtry, number of trees built (num.trees) and minimum number of nodes (min.node.size). This research aims to show the performance of a land valuation model using random forest with hyperparameter tuning analysis to get the best accuracy. The research results show that the land valuation model has an R^2 of 82.96% and a mean absolute percent error (MAPE) of 26.83, so the model has reasonable predictive ability. Based on the standard ratio issued by the IAAO, the land valuation model quality test gives a coefficient of variation (COV) value of 22.62, a coefficient of dispersion (COD) value of 25.29, and a price-related differential (PRD) value of 1.12. COV, COD, and PRD are still outside the specified tolerance limits.

Keywords : modeling, land valuation, random forest, hyperparameter

I. PENDAHULUAN

Nilai merupakan sebagai suatu opini dari manfaat ekonomi atas kepemilikan aset atau harga yang paling mungkin dibayarkan untuk suatu aset dalam pertukaran sehingga nilai bukan fakta. Nilai diperoleh melalui suatu proses yang disebut dengan penilaian. Penilaian didefinisikan sebagai proses pekerjaan seorang penilai dalam memberikan opini tertulis mengenai nilai ekonomi pada saat tertentu (KEPI dan SPI, 2018). Secara umum, penilaian dapat dibagi menjadi dua, yaitu penilaian individual yang dilaksanakan secara khusus terhadap satu objek penilaian dan penilaian massal yang dilaksanakan secara serempak dan sistematis terhadap sekelompok objek penilaian. Penilaian massal merupakan proses menilai sekelompok properti pada tanggal tertentu menggunakan sekumpulan data, metode standar, dan pengujian secara statistika dengan mengandalkan model atau persamaan penilaian melalui analisis matematis terhadap data pasar (IAAO, 2017). Penilaian massal memerlukan pembentukan model yang mampu mencerminkan kekuatan-kekuatan penawaran dan permintaan pada area yang luas pada sekelompok properti, bukan hanya pada satu properti. Model dapat diartikan sebagai perwujudan dari suatu abstraksi berbagai aspek realitas dunia nyata yang dibuat untuk satu atau berbagai tujuan (Insukindro, 1992). Model memang tidak sama dengan realitas, namun dari model yang baik akan dapat menerangkan dan meramalkan sebagian besar dari apa yang terjadi dengan realitas.

Model penilaian massal dihasilkan melalui pemodelan matematika. Kredibilitas model penilaian massal bergantung pada data yang digunakan dan keterampilan pembuat model. Pembuat model harus mengembangkan model yang digunakan dan memastikan model yang dibangun mencerminkan indikasi nilai pasar properti serta model tersebut dapat diadaptasi untuk berbagai kegunaan (KEPI dan SPI, 2018). Terdapat beberapa hal yang perlu diperhatikan dalam model penilaian massal antara lain model penilaian dibangun berdasarkan data yang sebanding, variabel yang digunakan bersifat rasional sehingga diperoleh hasil pemodelan yang wajar, terdapat konsistensi masing-masing properti yang tidak berubah secara tiba-tiba antar waktu dan model

penilaian dapat dijelaskan dengan mudah kepada pengguna nilai (Gloude-mans, 1982). Mengacu pada *standard on automated valuation models* (AVMs) yang dikeluarkan oleh International Association of Assessing Officers (IAAO), terdapat beberapa model penilaian massal antara lain regresi linear berganda, *artificial neural network*, dan *random forest* (IAAO, 2018).

Random forest pertama kali diperkenalkan oleh Breiman (2001), kini telah menjadi salah satu alat klasifikasi dan regresi nonparametrik untuk menghasilkan model prediksi berdasarkan variabel bebas tanpa membuat asumsi terlebih dahulu mengenai bentuk hubungan antara variabel bebas dengan variabel terikatnya (Probst et al., 2019). Saat ini *random forest* telah menjadi topik dalam beberapa ulasan dan penelitian. Salah satu fokus pengaplikasian *random forest*, yaitu pada konstruksi aturan klasifikasi atau regresi yang digunakan sebagai alat prediksi, di antaranya digunakan untuk prediksi harga properti atau nilai tanah.

Random forest memiliki tingkat akurasi yang baik sebagai alat prediksi (Putra et al., 2023). Bahkan sebagai alat prediksi, *random forest* memiliki tingkat akurasi yang paling baik di antara algoritma *machine learning* lainnya seperti *support vector machine*, *decision tree*, dan *artificial neural network* (Song et al., 2021). Karena memiliki kemampuan prediksi yang baik, *random forest* telah digunakan sebagai model untuk memprediksi harga properti maupun nilai tanah. Perbandingan harga prediksi properti dan harga aktualnya menunjukkan bahwa perhitungan harga prediksi properti menggunakan model *random forest* memiliki nilai prediksi yang dapat diterima (Adetunji et al., 2022). Dalam proses perhitungannya, *random forest* memerlukan pengaturan beberapa *hyperparameter* untuk mengontrol struktur pohon yang akan dibangun sehingga memberikan kemampuan prediksi yang baik (Nursalman & Mustikasari, 2020).

Penelitian ini bertujuan untuk menganalisis kemampuan algoritma *random forest* untuk memprediksi nilai tanah. Lokasi penelitian berada di Kota Surakarta dengan data nilai tanah yang telah dilakukan survei pada tahun 2021. Penulis menerapkan pembelajaran *random forest* dengan analisis pengaturan *hyperparameter* agar

mendapatkan model dengan hasil akurasi terbaik. Penelitian ini diharapkan dapat menambah literatur mengenai penilaian massal di Indonesia khususnya pemakaian metode kecerdasan buatan atau *artificial intelligence*.

II. METODE

A. Sumber Data, Populasi, dan Sampel

Lokasi penelitian berada di Kota Surakarta. Kota Surakarta, juga disebut Solo, adalah wilayah otonom dengan status kota di bawah Provinsi Jawa Tengah. Kota Surakarta memiliki luas 44,04 km² yang berbatasan dengan Kabupaten Karanganyar dan Kabupaten Boyolali di sebelah utara, Kabupaten Karanganyar dan Kabupaten Sukoharjo di sebelah timur dan barat, dan Kabupaten Sukoharjo di sebelah selatan. Secara administratif, Kota Surakarta terbagi ke dalam 5 kecamatan yaitu Kecamatan Pasar Kliwon, Kecamatan Jebres, Kecamatan Banjarsari, Kecamatan Laweyan, dan Kecamatan Serengan.

Penelitian ini menggunakan pendekatan kuantitatif. Pendekatan kuantitatif ialah pendekatan yang di dalam usulan penelitian, proses, hipotesis, turun ke lapangan, analisis data, dan kesimpulan data sampai dengan penulisannya mempergunakan aspek pengukuran, perhitungan, rumus, dan kepastian data numerik (Musianto, 2002). Data kuantitatif dalam penelitian ini diperoleh dari hasil perhitungan data spasial maupun data nonspasial. Bahan yang digunakan dalam penelitian ini terdiri dari data spasial dan data nonspasial.

1. Data spasial yang digunakan adalah
 - a. peta bidang tanah (PBT) tahun 2021 yang berjumlah sekitar 134.892 bidang tanah yang diperoleh dari Kementerian Agraria dan Tata Ruang/Badan Pertanahan Nasional (ATR/BPN);
 - b. peta batas administrasi yang diperoleh dari Badan Informasi Geospasial (BIG);
 - c. peta jaringan jalan yang diperoleh dari Badan Informasi Geospasial (BIG); dan
 - d. peta rencana tata ruang wilayah (RTRW) Tahun 2011-2031 yang diperoleh dari Badan Perencanaan Pembangunan Daerah (Bappeda) Kota Surakarta.

2. Data nonspasial yang digunakan adalah:
 - a. data transaksi jual beli properti selama kurun waktu tahun 2020-2021 yang diperoleh dari hasil survei; dan
 - b. data penawaran properti tahun 2021 yang diperoleh dari hasil survei.

Data transaksi dan penawaran tersebut merupakan data hasil kegiatan pengumpulan data nilai tanah yang telah dilakukan oleh Direktorat Penilaian Tanah dan Ekonomi Pertanahan, Kementerian ATR/BPN. Populasi dalam penelitian ini ialah seluruh bidang tanah di Kota Surakarta pada tahun 2021 baik yang sudah terdaftar suatu hak maupun yang belum terdaftar. Sampel dalam penelitian ini merupakan bidang tanah yang telah terjadi transaksi jual beli maupun penawaran sebagaimana dijelaskan pada data nonspasial.

B. Variabel Penilaian Tanah

Sebagai barang ekonomis, tanah mempunyai kekuatan ekonomi di mana nilai tanah atau harga tanah sangat dipengaruhi oleh permintaan dan penawaran. Dalam kurva penawaran dan permintaan, kurva penawaran tanah dalam jangka pendek sangat bersifat in-elastis yang berarti bahwa nilai tanah atau harga tanah sangat bergantung pada pergerakan kurva permintaan. Naiknya kurva permintaan tanah dapat mengakibatkan kenaikan nilai tanah atau harga tanah yang cukup signifikan.

Selain permintaan dan penawaran, nilai tanah dipengaruhi oleh faktor lain seperti faktor sosial, faktor pemerintah, dan faktor fisik. Faktor-faktor yang memengaruhi nilai tanah tersebut kemudian diterjemahkan sebagai variabel penilaian tanah. Dalam penelitian ini digunakan variabel penilaian tanah yang dapat diukur secara kuantitatif yaitu variabel dari faktor fisik seperti lebar depan, bentuk, letak, lebar jalan, kelas jalan, dan variabel dari faktor pemerintah seperti *zoning* (peruntukan kawasan), penyediaan fasilitas, dan pelayanan oleh pemerintah.

Variabel yang digunakan dalam model penilaian tanah dibagi menjadi dua variabel, yaitu variabel terikat dan variabel bebas. Variabel terikat yaitu nilai tanah sedangkan variabel bebas merupakan faktor-faktor yang dianggap memengaruhi dengan nilai tanah. Variabel dan tipe data dijelaskan pada Tabel 1.

Tabel 1 Variabel Penilaian Tanah Massal

No.	Variabel	Tipe Data	Keterangan	Perolehan Data
1.	Nilai tanah	Rasio	Nilai tanah yang dinyatakan dalam satuan moneter rupiah	Survei lapangan
2.	Lebar depan	Rasio	Lebar depan sisi bidang tanah yang menghadap ke akses jalan terdekat	Pengukuran pada peta bidang tanah (<i>spatial analysis</i>)
3.	Bentuk	Ordinal	Bentuk bidang tanah 1) tidak beraturan 2) trapesium 3) persegi	Pengamatan pada peta bidang tanah (<i>spatial analysis</i>)
4.	Zonasi	Ordinal	Zonasi atau penggunaan tanah 1) lindung dan sempadan 2) pertanian 3) permukiman sederhana 4) permukiman menengah 5) permukiman mewah 6) industri 7) perdagangan dan jasa	Analisis spasial pada peta bidang tanah dengan peta rencana tata ruang wilayah (<i>spatial analysis</i>)
5.	Letak	Ordinal	Posisi bidang tanah terhadap jalan terdekat 1) lain-lain (tidak memiliki akses) 2) tusuk sate 3) normal 4) <i>hook</i>	Analisis spasial pada peta bidang tanah dengan peta jaringan jalan (<i>spatial analysis</i>)
6.	Kelas jalan	Ordinal	Kelas jalan terdekat bidang tanah 1) setapak 2) lokal sekunder 3) lokal primer 4) kolektor sekunder 5) kolektor primer 6) arteri sekunder 7) arteri primer	Analisis spasial pada peta jaringan jalan (<i>spatial analysis</i>)
7.	Lebar jalan	Rasio	Lebar jalan sebagai akses utama bidang tanah	Analisis spasial pada peta jaringan jalan (<i>spatial analysis</i>)
8.	Jarak bidang tanah terhadap jalan utama	Rasio	Jarak bidang tanah yang diukur berdasarkan jaringan jalan terhadap jalan utama yaitu arteri primer, arteri sekunder, kolektor primer dan kolektor sekunder	Analisis spasial jarak terdekat antara bidang tanah terhadap jalan utama berdasarkan jaringan jalan (<i>network analysis</i>)
9.	Jarak bidang tanah terhadap fasilitas	Rasio	Jarak bidang tanah yang diukur berdasarkan jaringan jalan terhadap fasilitas umum terdekat yaitu pusat ekonomi, keraton, fasilitas kesehatan, fasilitas transportasi, dan fasilitas pendidikan	Analisis spasial jarak terdekat antara bidang tanah terhadap fasilitas berdasarkan jaringan jalan (<i>network analysis</i>)
10.	Risiko	Nominal	Faktor risiko yang terjadi di lokasi penelitian seperti banjir	Analisis spasial (<i>spatial analysis</i>)

Sumber: Direktorat Penilaian Tanah dan Ekonomi Pertanian, 2021

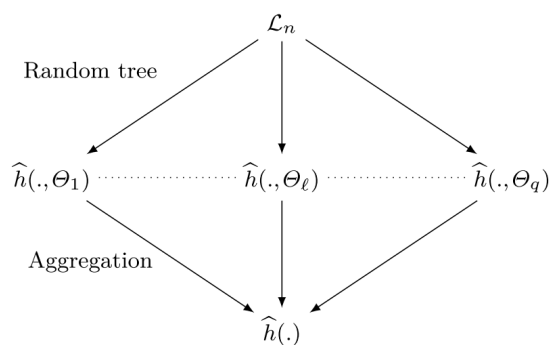
Variabel bentuk, zonasi, letak, dan kelas jalan merupakan data kualitatif yang diubah menjadi data kuantitatif melalui metode *scoring*. Makin tinggi skor bermakna jika kondisi tersebut secara persepsi lebih diinginkan daripada yang memiliki skor rendah. Sebagai contoh variabel letak memiliki skor (1) tidak memiliki akses (2) tusuk sate (3) normal (4) *hook*, secara persepsi bidang tanah yang memiliki akses jalan pada kedua sisi (*hook*) lebih diinginkan daripada hanya memiliki akses jalan pada salah satu

sisi saja (normal). Begitu juga bidang tanah yang memiliki akses pada salah satu sisi bidang (normal) lebih diinginkan daripada bidang tanah yang terletak pada tusuk sate.

C. Random Forest

Random forest adalah kombinasi dari pohon-pohon prediktor (*trees*) sedemikian rupa sehingga setiap pohon bergantung pada nilai acak (*random*) yang diambil sampelnya secara mandiri dan dengan

distribusi yang sama untuk semua pohon di hutan (*forests*). Metode *random forest* merupakan pengembangan dari metode *classification and regression tree* (CART) dengan menerapkan metode *bootstrap aggregating (bagging)* dan *random feature selection* (Breiman, 2001). Prinsip umum dari *random forest* adalah mengumpulkan kumpulan pohon keputusan acak dengan tujuan berupaya mengoptimalkan sebuah prediktor sekaligus seperti pada pohon CART. Secara umum skema *random forest* sebagaimana pada Gambar 1.



Sumber: Genuer Robin & Poggi Jean-Michel, 2020

Gambar 1 Skema Umum *Random Forest*

Misal $(\hat{h}(\cdot, \theta_1), \dots, \hat{h}(\cdot, \theta_q))$ sebagai kumpulan pohon prediktor (*trees*), dengan $\theta_1, \dots, \theta_q$ sebagai variabel bebas acak dari \mathcal{L}_n . Prediktor *random forest* \hat{h}_{RF} diperoleh dengan menggabungkan kumpulan pohon acak (*random trees*). Agregasi dilakukan sebagai berikut.

1. (rata-rata dari prediksi pohon tunggal) dalam regresi $\hat{h}_{RF}(x) = \frac{1}{q} \sum_{\ell=1}^q \hat{h}(x, \theta_{\ell})$ (1)
2. (suara mayoritas di antara prediksi pohon tunggal) $\hat{h}_{RF}(x) = \underset{1 \leq c \leq C}{arg\ max} \frac{1}{q} \sum_{\ell=1}^q 1_{\hat{h}(x, \theta_{\ell})=c}$ (2)

Breiman (2001) menyatakan bahwa dengan menggabungkan prediksi dari beberapa CART independen dan menambahkan *bagging* dalam prosesnya dapat meningkatkan kinerja dan generalisasi model menjadi lebih baik.

Algoritma *random forest* bekerja dengan sejumlah *hyperparameter*. Kinerja algoritma *random forest* dapat ditingkatkan dengan mengatur *hyperparameter*. Sebagaimana diuraikan dalam Breiman (2001), keacakan yang digunakan dalam membangun konstruksi pohon (*trees*) harus bertujuan untuk menghasilkan korelasi ρ yang rendah dengan

tetap mempertahankan kekuatan yang wajar. Hal ini dikenal sebagai pertukaran (*trade-off*) antara bias dan varian. Kondisi tersebut dapat dikontrol dengan *hyperparameter* antara lain *mtry*, ukuran sampel, dan ukuran *node* (Probst et al., 2019).

1. Jumlah kandidat variabel yang diambil secara acak (*mtry*)

Salah satu *hyperparameter* utama *random forest* adalah *mtry* yang didefinisikan sebagai jumlah kandidat variabel yang diambil dan dipilih secara acak pada setiap *node* ketika membangun pohon (*trees*). Nilai *mtry* yang lebih rendah akan menghasilkan konstruksi antar pohon yang lebih berbeda dan memiliki korelasi lebih sedikit sehingga menghasilkan stabilitas yang lebih baik ketika digabungkan. Beberapa perangkat lunak mengatur nilai *mtry* dengan nilai \sqrt{p} untuk klasifikasi dan $p/3$ untuk regresi dengan p sebagai jumlah variabel bebas (Genuer Robin & Poggi Jean-Michel, 2020).

2. Jumlah pohon yang dibangun (*num.trees*)

Beberapa perangkat lunak mengatur jumlah pohon (*num.trees*) dengan nilai bawaan 500. Pada pengolahan dengan jumlah data yang relatif besar, peningkatan kinerja *random forest* dapat dicapai dengan membangun jumlah pohon minimal 100 (Probst et al., 2019).

3. Jumlah minimal observasi pada setiap *node* (*min.node.size*)

Jumlah minimum observasi yang harus dimiliki oleh sebuah *node* secara bawaan adalah 1 untuk klasifikasi dan 5 untuk regresi. Nilai tersebut dipercaya memberikan hasil kinerja *random forest* yang bagus (Díaz-Uriarte & Alvarez de Andrés, 2006; Goldstein et al., 2011; Probst et al., 2019).

Proses perhitungan dan analisis data model penilaian menggunakan algoritma *random forest* pada penelitian ini menggunakan paket "*Ranger*" dan paket "*Caret*" untuk mengatur *hyperparameter* agar diperoleh model *random forest* yang optimal. Paket "*Ranger*" merupakan paket program yang dibangun pada bahasa R sebagai implementasi algoritma dari *random forest* (Breiman, 2001) yang dapat diproses secara cepat terutama cocok untuk data berdimensi tinggi (Wright et al., 2023). Paket ini berfungsi untuk

melakukan perhitungan klasifikasi dan regresi menggunakan algoritma *random forest* berdasarkan data masukan yang digunakan. Paket “*Caret*” (kependekan dari *classification and regression training*) merupakan paket program yang dibangun pada bahasa R sebagai serangkaian fungsi yang berupaya menyederhanakan proses pembuatan model prediktif (Kuhn et al., 2023). Paket “*Caret*” menyediakan antarmuka terpadu untuk melakukan tugas *machine learning*, seperti prapemrosesan data, pelatihan model, dan evaluasi kinerja.

D. Uji Kualitas Model Penilaian Tanah

Uji kualitas model penilaian tanah dilakukan dengan melihat besarnya nilai *coefficient of variation* (COV), *coefficient of dispersion* (COD) dan *price related differential* (PRD) rasio antara hasil perhitungan model atau nilai tanah prediksi terhadap nilai tanah aktual (IAAO, 2013).

1. *Coefficient of Dispersion* (COD)

COD merupakan salah satu ukuran yang digunakan untuk mengukur tingkat keseragaman penilaian. COD mengukur rata-rata penyimpangan sekelompok angka dari median dan dinyatakan dalam persentase median. Penafsiran nilai COD tidak bergantung pada asumsi bahwa data terdistribusi secara normal. Nilai COD yang disarankan < 20%.

2. *Coefficient of Variation* (COV)

COV menunjukkan tingkat akurasi model dalam memprediksi nilai tanah. COV merupakan ukuran statistik standar dari penyebaran relatif data sampel terhadap rata-rata data atau standar deviasi yang dinyatakan sebagai persentase mean. Nilai COV yang disarankan < 20%.

3. *Price Related Differential* (PRD)

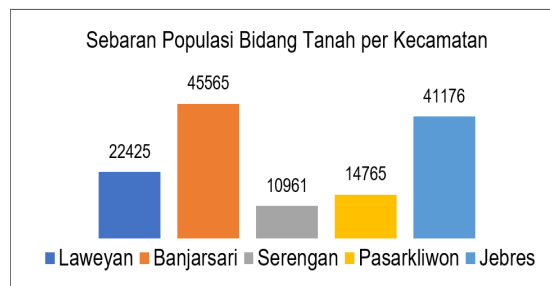
PRD mengukur rata-rata dibagi dengan rata-rata tertimbang. Nilai PRD digunakan untuk mengukur tingkat kewajaran prediksi nilai tanah. Nilai PRD yang baik seharusnya mendekati 1,00. Nilai PRD hendaknya berkisar antara 0,98 sampai dengan 1,03. Jika nilai PRD kurang dari 0,98 dikatakan telah terjadi progresivitas yang bermakna bahwa prediksi nilai model berada di atas nilai sebenarnya dan jika lebih dari

1,03 maka terjadi regresivitas yang bermakna bahwa prediksi nilai model berada di bawah nilai sebenarnya.

III. HASIL DAN PEMBAHASAN

A. Dataset Sampel

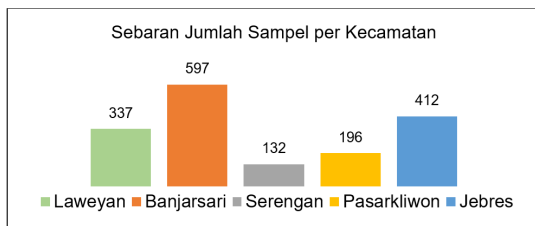
Populasi dalam penelitian ini ialah seluruh bidang tanah di Kota Surakarta pada tahun 2021 yang berjumlah kurang lebih 134.982 bidang tanah. Dari seluruh bidang tanah tersebut, kurang lebih 90% atau 122.699 bidang tanah sudah terdaftar dengan suatu hak atas tanah sedangkan kurang lebih 9% atau 12.193 bidang tanah belum terdaftar dengan suatu hak tanah. Bidang tanah yang sudah terdaftar terdiri dari 86,9% hak milik, 3% hak guna bangunan, 0,7% hak pakai, dan 0,3% tanah wakaf. Apabila diamati sebaran populasi bidang tanah setiap kecamatan maka jumlah populasi bidang tanah paling banyak berada di Kecamatan Banjarsari sekitar 34% dari seluruh bidang tanah, sedangkan jumlah populasi bidang tanah paling sedikit berada di Kecamatan Serengan sekitar 8% dari seluruh bidang tanah. Sebaran populasi bidang tanah setiap kecamatan dapat dilihat pada Gambar 2 berikut.



Sumber: Olah data, 2024

Gambar 2 Sebaran Populasi Bidang Tanah Setiap Kecamatan

Sampel dalam penelitian ini merupakan bidang tanah yang telah terjadi transaksi jual beli maupun penawaran yang berjumlah 1.651 sampel nilai tanah yang tersebar di Kota Surakarta. Sebaran sampel nilai tanah terbanyak berada di Kecamatan Banjarsari yang memang secara administratif memiliki luas wilayah yang paling luas sedangkan sampel nilai tanah paling sedikit berada di Kecamatan Serengan. Jika diamati proporsi sebaran sampel setiap kecamatan berbanding lurus dengan jumlah populasi bidang tanahnya. Sebaran sampel nilai tanah setiap kecamatan dapat dilihat pada Gambar 3.



Sumber: Olah data, 2024

Gambar 3 Sebaran Sampel Nilai Tanah Setiap Kecamatan

Sebelum dilakukan pengolahan data menggunakan algoritma *random forest* maka data dibagi menjadi data *training* dan data *testing*. Data *training* digunakan untuk melatih algoritma *random forest* dalam membangun model penilaian tanah sedangkan data *testing* digunakan untuk menguji kemampuan model. Data *training* dan data *testing* dapat dibagi dengan perbandingan 70:30, 80:20 atau 90:10 (Muningsih, 2022). Penelitian ini menggunakan perbandingan data sebanyak 80% untuk data *training* dan 20% untuk data *testing* yang dibagi secara *random*. Perbandingan data setiap kecamatan dapat diamati pada Tabel 2 berikut.

Tabel 2 Jumlah Data *Training* dan Data *Testing*

No.	Kecamatan	Jumlah data <i>training</i>	Jumlah data <i>testing</i>
1	Banjarsari	490	100
2	Jebres	328	80
3	Laweyan	263	69
4	Pasarkliwon	143	48
5	Serengan	99	31
Total		1.323	328

Sumber: Olah data, 2024

B. Random Forest untuk Model Penilaian Tanah Massal

Algoritma *random forest* dapat diaplikasikan untuk klasifikasi maupun regresi. *Random forest* dengan fungsi klasifikasi dihasilkan ketika variabel terikat berupa data kategori sedangkan fungsi regresi dihasilkan ketika variabel terikat berupa data numerik (Suliztia, 2020). Model 1 penilaian tanah dalam penelitian ini menggunakan fungsi regresi dengan persamaan sebagai berikut.

$$\text{Nilai} = \text{lb_dpn} + \text{s_bentuk} + \text{s_zonasi} + \text{s_letak} + \text{s_kls_jln} + \text{lb_jln} + \text{jk_atrp} + \text{jk_atrs} + \text{jk_kolp} + \text{jk_kols} + \text{CBD} + \text{Keraton} + \text{RS} + \text{Trans} + \text{Univ} + \text{Genanga} \quad (3)$$

keterangan:

nilai : nilai tanah

lb_dpn : lebar depan

s_bentuk : bentuk

s_letak : letak

s_kls_jln : kelas jalan

lb_jln : lebar jalan

jk_atrp : jarak ke jalan arteri primer

jk_atrs : jarak ke jalan arteri sekunder

jk_kolp : jarak ke jalan kolektor primer

jk_kols : jarak ke jalan kolektor

CBD : jarak ke CBD

Keraton : jarak ke keraton

RS : jarak ke rumah sakit

Trans : jarak ke transportasi

Univ : jarak ke universitas

Genanga : faktor banjir

Untuk memperluas model penilaian tanah, dengan persamaan fungsi regresi yang sama pada persamaan (3) dilakukan transformasi data variabel jarak ke jalan arteri primer, jarak ke jalan arteri sekunder, jarak ke jalan kolektor primer, jarak ke jalan kolektor sekunder, jarak ke CBD, jarak ke keraton, jarak ke rumah sakit, jarak ke transportasi, jarak ke universitas ke dalam bentuk logaritma natural. Transformasi data ini bertujuan untuk menyamakan variasi data karena adanya perbedaan satuan. Model 2 penilaian tanah menggunakan fungsi regresi dengan transformasi data ke dalam bentuk logaritma natural ditunjukkan pada persamaan sebagai berikut.

$$\text{Nilai} = \text{lb_dpn} + \text{s_bentuk} + \text{s_zonasi} + \text{s_letak} + \text{s_kls_jln} + \text{lb_jln} + \ln_jk_atrp + \ln_jk_atrs + \ln_jk_kolp + \ln_jk_kols + \ln_CBD + \ln_Keraton + \ln_RS + \ln_Trans + \ln_Univ + \text{Genanga} \quad (4)$$

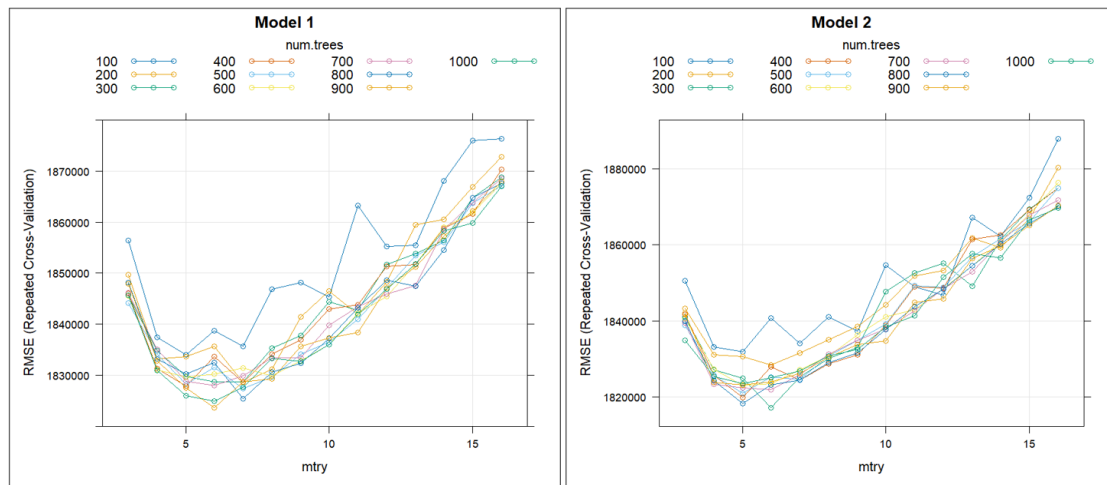
Keterangan: \ln_jk_atrp , \ln_jk_atrs , \ln_jk_kolp , \ln_jk_kols , \ln_CBD , $\ln_Keraton$, \ln_RS , \ln_Trans , \ln_Univ merupakan tranformasi data ke dalam bentuk logaritma natural pada variabel-variabel jarak.

Algoritma *random forest* bekerja dengan sejumlah *hyperparameter*. Kinerja algoritma *random forest* dapat ditingkatkan dengan mengatur *hyperparameter*. *Hyperparameter* yang diatur antara lain *min.node.size* atau jumlah minimum observasi pada setiap node, *num.trees* atau jumlah pohon

(trees) yang dibangun dan *mtry* atau jumlah kandidat variabel yang dipilih secara acak.

Jumlah minimum observasi pada setiap *node* atau *min.node.size* yaitu 5 untuk regresi dan 1 untuk klasifikasi dipercaya memberikan hasil kinerja *random forest* yang bagus (Díaz-Uriarte & Alvarez de Andrés, 2006; Goldstein et al., 2011; Probst et al., 2019). Pengaturan *mtry* dan *num.trees* dilakukan dengan cara melakukan pelatihan terhadap data

training secara berulang dengan mengombinasikan pengaturan *hyperparameter mtry* dan *num.trees*. Nilai *hyperparameter mtry* diatur dengan nilai 3 sampai 16 sedangkan *hyperparameter num.trees* diatur dengan nilai 100, 200, 300, 400, 500, 600, 700, 800, 900, dan 1000. Model dengan pengaturan *hyperparameter* terbaik akan memberikan nilai *root mean square error* (RMSE) terkecil. Hasil perhitungan *mtry*, *num.trees* dan RMSE menggunakan *r package "Ranger"* dan "*Caret*" dapat diamati pada Gambar 4 berikut.



Sumber: Olah data, 2024.

Gambar 4 *mtry*, *num.trees* dan RMSE pada Model 1 dan Model 2

Gambar 4 menunjukkan bahwa pengaturan nilai *mtry* antara 6 sampai 9 pada model 1 maupun model 2 penilaian tanah relatif memberikan nilai RMSE yang kecil. Makin besar nilai *mtry* maka nilai RMSE yang dihasilkan juga cenderung makin besar. Pengaturan *hyperparameter* pada model 1 diperoleh hasil nilai RMSE terkecil 1823602 dengan nilai optimal *mtry* sebesar 6 dan *num.trees* sebesar 900, sedangkan pada model 2 diperoleh nilai RMSE terkecil 1817228 dengan nilai optimal *mtry* sebesar 6 dengan *num.trees* sebesar 300. Selanjutnya, nilai *mtry* dan *num.trees* tersebut digunakan sebagai *hyperparameter* untuk membuat model penilaian tanah dengan algoritma *random forest* terhadap data *training*. Kinerja model penilaian tanah ditunjukkan pada Tabel 3.

Tabel 3 Kinerja Model Penilaian Tanah

<i>Hyperparameter</i>	Model 1	Model 2
<i>mtry</i>	6	6
<i>num.trees</i>	900	300
<i>min.node.size</i>	5	5
R ² (OOB)	0,8296	0,8278

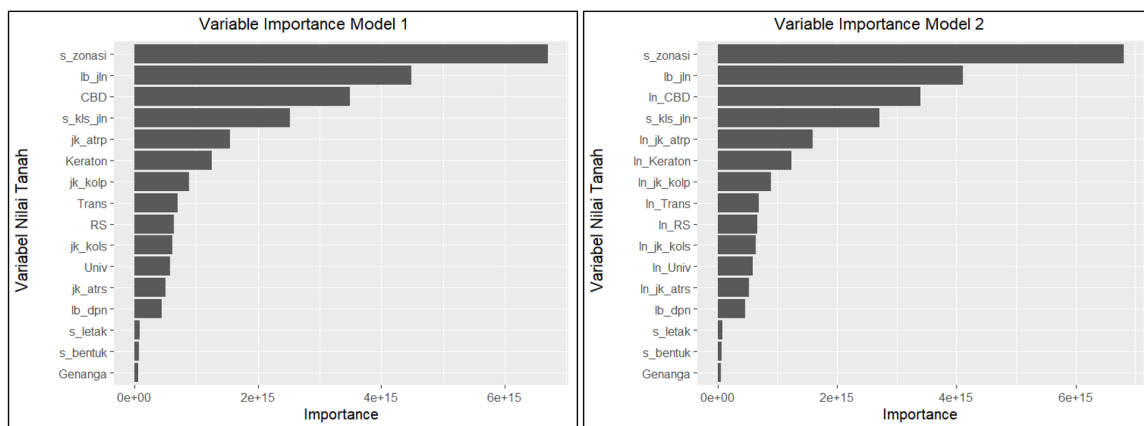
Sumber: Olah data, 2024

Model 1 memiliki nilai R² sebesar 82,96% yang berarti bahwa variabel bebas nilai tanah dapat menjelaskan variabel terikat nilai tanah sebesar 82,96% sedangkan sisanya sebesar 17,04% dijelaskan oleh variabel bebas di luar model 1. Begitu juga dengan model 2 memiliki nilai R² sebesar 82,78% yang berarti bahwa variabel bebas nilai tanah dapat menjelaskan variabel terikat nilai

tanah sebesar 82,78% sedangkan sisanya sebesar 17,22% dijelaskan oleh variabel bebas di luar model 2. Tampaknya baik model 1 ataupun model 2 memiliki perbedaan nilai R^2 yang tidak signifikan. Hal ini mengindikasikan model 1 dan model 2 memiliki kemampuan atau kinerja yang sama.

Random forest menghasilkan ukuran tingkat kepentingan (*variable importance*) variabel bebas untuk mengidentifikasi variabel bebas yang relevan dengan variabel terikat. Ukuran prioritas variabel

bebas dapat diketahui melalui ukuran tingkat kepentingan (*variable importance*) variabel bebas. Ukuran tingkat kepentingan (*variable importance*) merupakan perhitungan berapa kali variabel bebas muncul dalam kelompok pohon keputusan sehingga makin besar dampaknya, makin penting variabel tersebut (Christy & Suryowati, 2021). Ukuran tingkat variabel kepentingan nilai tanah pada model 1 dan model 2 penilaian massal dapat diamati pada Gambar 5 berikut.



Sumber: olah data, 2024

Gambar 5 Ukuran Tingkat Kepentingan Variabel Nilai Tanah

Baik model 1 maupun model 2 memiliki prioritas ukuran tingkat kepentingan variabel yang sama. Variabel zonasi atau penggunaan tanah, jalan dan jarak terhadap pusat perdagangan memiliki tingkat kepentingan paling besar terhadap nilai tanah. Sejalan dengan Teori Von Thunen yang menjelaskan bahwa nilai sewa tanah dipengaruhi oleh pola penggunaan tanah dan aksesibilitas atau jaraknya terhadap pusat kota atau pasar. Variabel letak bidang tanah, bentuk bidang tanah dan faktor risiko banjir memiliki tingkat kepentingan yang paling kecil terhadap nilai tanah.

Tabel 4 Nilai MAPE Model Penilaian Tanah

Model	Data Training	Data Testing
Model 1	25,76	26,83
Model 2	25,80	27,00

Sumber: penulis, 2024

Untuk menguji kemampuan prediksi model terhadap data baru maka digunakan data *testing* yang telah dibagi pada *dataset* sampel sebelumnya.

Sebagaimana pada Tabel 4, model 1 menghasilkan nilai MAPE (*mean absolute percentage error*) yang sedikit lebih kecil daripada model 2, walaupun perbedaan nilai MAPE tersebut terlihat tidak terlalu signifikan. Model 1 memiliki kemampuan prediksi yang lebih baik daripada Model 2. Nilai MAPE kurang dari 30% menunjukkan bahwa model penilaian tanah memiliki kemampuan prediksi yang masih layak (Montaño Moreno et al., 2013).

C. Uji Kualitas Model Penilaian Tanah

Berdasarkan standar yang diterbitkan oleh IAAO, uji kualitas model penilaian tanah massal dapat dilakukan dengan melihat nilai COV, COD, dan PRD rasio antara nilai tanah prediksi atau hasil perhitungan model terhadap nilai tanah aktual. Sebagaimana Tabel 5, hasil perhitungan nilai COV, COD, dan PRD baik untuk model 1 maupun model 2 dengan data *training* ataupun data *testing* masih melebihi toleransi yang ditetapkan.

Tabel 5 Nilai COV, COD, dan PRD Model Penilaian Tanah

Model	COV		COD		PRD	
	Training	Testing	Training	Testing	Training	Testing
Model 1	23,20	22,62	24,57	25,29	1,11	1,12
Model 2	23,53	22,54	24,60	24,98	1,11	1,12

Sumber: Olah data, 2024

Nilai COV dan COD lebih dari 20% yang berarti model penilaian tanah yang dibangun belum mampu menangkap dengan baik fenomena nilai tanah di Kota Surakarta. Nilai PRD sebesar 1,11 masih di atas batas yang ditetapkan sebesar 0,98 – 1,03 yang berarti bahwa terjadi regresivitas yaitu prediksi nilai model berada di bawah nilai aktualnya.

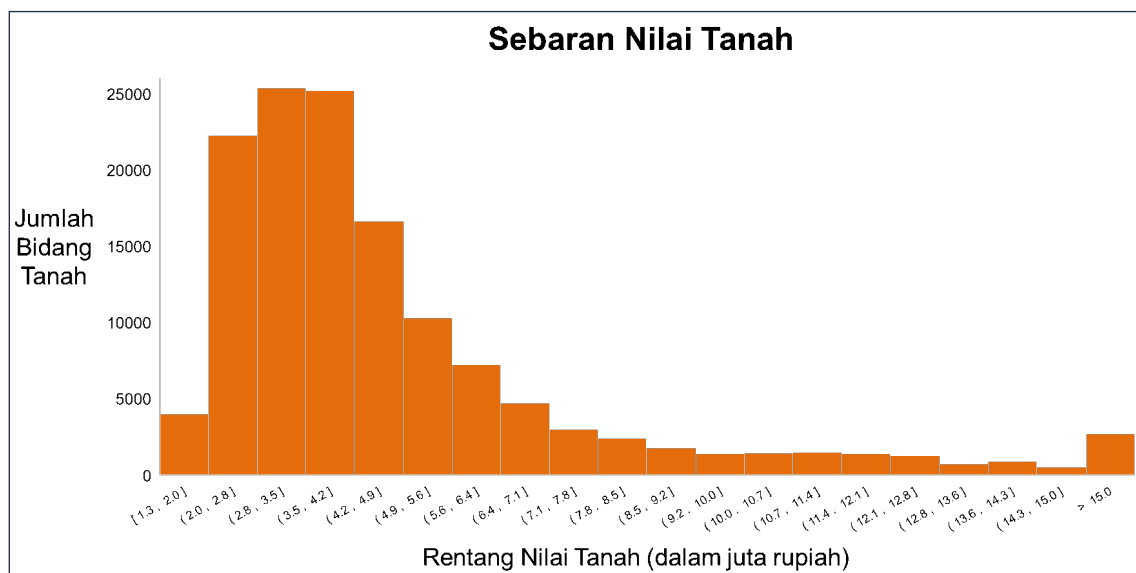
Isu heterogenitas spasial harus dipertimbangkan ketika mencoba memahami model nilai tanah sehingga dapat memberikan pemahaman yang lebih mendalam mengenai terjadinya anomali nilai tanah di suatu wilayah (Páez et al., 2001; Ren et al., 2021; Wooseok et al., 2022). Kecenderungan terjadinya heterogenitas spasial pada nilai tanah mengakibatkan model penilaian tanah yang telah dibangun kurang mampu memahami fenomena anomali nilai tanah antarlokasi. Heterogenitas spasial menyebabkan variabel bebas yang sama akan memberikan respon yang berbeda antarlokasi yang berbeda.

Saran untuk penelitian selanjutnya dalam membangun model penilaian tanah agar

memperhatikan aspek heterogenitas spasial nilai tanah. Model yang dibangun tidak harus melingkupi satu kota/kabupaten sekaligus tetapi, dapat dibagi ke dalam satuan wilayah yang lebih kecil misal per kecamatan. Selain itu, dapat menggunakan model yang memperhatikan aspek heterogenitas spasial.

D. Peta Nilai Tanah Berbasis Bidang Tanah

Model penilaian tanah yang telah dibangun selanjutnya digunakan untuk menghitung nilai prediksi seluruh bidang tanah yang ada di Kota Surakarta. Pada bab ini akan digunakan Model 1 penilaian tanah untuk memprediksi seluruh bidang tanah yang berjumlah sekitar 134.892 bidang tanah. Hasil prediksi nilai tanah paling rendah sebesar Rp1.312.230,- sedangkan nilai tanah paling tinggi sebesar Rp29.422.280,-. Sebaran nilai tanah ditampilkan pada Gambar 6 menunjukkan nilai tanah paling banyak terdapat pada rentang nilai 2,0 juta sampai 4,2 juta.



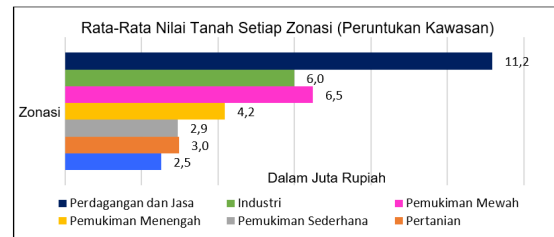
Sumber: olah data, 2024

Gambar 6 Sebaran Nilai Tanah di Kota Surakarta

Rata-rata nilai tanah setiap zonasi dapat diamati pada Gambar 7. Bidang tanah yang terletak pada zonasi perdagangan dan jasa memiliki rata-rata nilai tanah yang paling tinggi, bahkan hampir dua kali lipat dari rata-rata nilai tanah pada zonasi industri atau permukiman mewah. Properti yang dibangun pada zonasi perdagangan dan jasa dapat digunakan untuk usaha komersial sehingga membentuk ekspektasi bahwa properti tersebut akan memberikan keuntungan atau pendapatan kepada pemiliknya pada masa yang akan datang. Selain itu, kawasan perdagangan dan jasa memiliki aksesibilitas yang sangat baik didukung dengan jalan yang lebar dan tersedianya fasilitas di sekitarnya. Keuntungan inilah yang mendorong tingginya nilai tanah pada kawasan perdagangan dan jasa.

Zonasi permukiman mewah memiliki rata-rata nilai tanah yang cukup tinggi di bawah zonasi perdagangan dan jasa. Kawasan permukiman mewah cenderung memiliki kondisi lingkungan yang nyaman. Aksesibilitas yang mudah dijangkau dan fasilitas-fasilitas umum yang tersebar di sekitarnya. Seseorang akan memberikan penawaran nilai tanah yang tinggi demi kenyamanan dan kemudahan aksesibilitas tersebut.

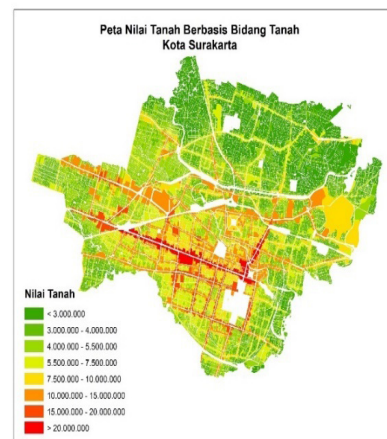
Apabila diamati kembali Gambar 7 terlihat bahwa rata-rata nilai tanah pada zonasi permukiman sederhana hampir sama bahkan lebih rendah dari rata-rata nilai tanah pada zonasi pertanian. Di Kota Surakarta, zonasi pertanian cenderung berada pada wilayah sebelah utara barat. Wilayah ini merupakan wilayah pengembangan perumahan. Walaupun kawasan tersebut merupakan zonasi pertanian tetapi sudah mulai masif dilakukan alih fungsi tanah pertanian menjadi tanah nonpertanian. Maraknya alih fungsi lahan akan mendorong naiknya nilai tanah sekalipun dalam wilayah tersebut masih zonasi pertanian. Ditambah regulasi kebijakan mengenai Rencana Tata Ruang Wilayah Kota Surakarta tahun 2021-2041 bahwa Kota Surakarta sudah tidak terdapat lagi zona peruntukan lahan pertanian. Selain itu, wilayah Kota Surakarta sebelah utara barat dilalui jalan utama menuju Kartasura dan Colomadu serta Bandara Adi Sumarmo mendorong tumbuhnya kawasan permukiman di wilayah tersebut, yang akhirnya akan mendorong permintaan tanah sehingga nilai tanah akan naik.



Sumber: Olah data, 2024

Gambar 7 Rata-Rata Nilai Tanah Setiap Zona (Peruntukan Kawasan)

Hasil prediksi nilai tanah setiap bidang tanah dapat digambarkan pada peta bidang tanah sehingga dapat diamati sebaran nilainya dalam bentuk spasial atau peta. Sebaran nilai tanah dengan nilai relatif tinggi berada di Jalan Slamet Riyadi. Kawasan di sekitar Jalan Slamet Riyadi merupakan kawasan pusat perdagangan dan jasa di Kota Surakarta. Kawasan tersebut juga berada dekat dengan Keraton Kasunanan Surakarta sebagai *landmark* dan kawasan wisata Kota Surakarta. Selain itu, kawasan Jalan Slamet Riyadi juga memiliki daya tarik yang besar karena jalan ini sebagai pintu masuk dan koridor utama Surakarta menuju pusat kota sekaligus memiliki banyak persimpangan yang menghubungkan dengan pusat kegiatan perekonomian di sekitarnya.



Sumber: Olah data, 2024.

Gambar 8 Peta Nilai Tanah Berbasis Bidang Tanah Kota Surakarta

Sebaran nilai tanah dengan nilai relatif rendah cenderung berada di kawasan sebelah utara Kota Surakarta terutama di Kecamatan Banjarsari dan

Kecamatan Jebres. Secara aksesibilitas, kawasan sebelah utara relatif jauh dari pusat kegiatan perekonomian di Kota Surakarta. Kondisi di kawasan utara Kota Surakarta juga masih memiliki beberapa permasalahan seperti kondisi jalan yang rusak dan masih sedikitnya pembangunan sehingga mengakibatkan ketidaktertarikan untuk penanaman modal dan pengembangan di kawasan utara Surakarta (Pamungkas et al., 2019). Hal tersebut secara ekonomi kurang menjadi faktor pendorong meningkatnya permintaan dan penawaran tanah atau properti sehingga nilai tanah cenderung lebih rendah. Pengembangan wilayah Kota Surakarta dan sekitarnya cenderung ke arah selatan di daerah Solo Baru dan ke arah barat di daerah Kartasura dan Colomadu.

IV. KESIMPULAN

Pengaturan *hyperparameter random forest* untuk model 1 penilaian tanah diperoleh nilai *mtry* sebesar 6, nilai *num.trees* sebesar 900 dan *min.node.size* sebesar 5 sedangkan untuk model 2 penilaian tanah diperoleh nilai *mtry* sebesar 6, nilai *num.trees* sebesar 300 dan *min.node.size* sebesar 5. Model penilaian tanah massal memiliki kemampuan prediksi yang cukup layak dibuktikan dengan nilai R^2 yang dihasilkan model 1 sebesar 82,96% sedangkan model 2 sekitar 82,78%, begitu juga nilai MAPE baik model 1 dan model 2 untuk data *training* maupun data *testing* masih di bawah 30%.

Uji kualitas model penilaian tanah menghasilkan nilai COV) nilai COD lebih dari 20% dan nilai PRD di luar rentang 0,98 - 1,02. Nilai COV, COD dan PRD tersebut masih di luar batas toleransi yang ditetapkan oleh IAAO. Hal ini diduga karena adanya heterogenitas spasial yang cenderung terjadi pada fenomena nilai tanah. Untuk menanggulangi hal tersebut dapat ditempuh dengan membagi model penilaian tanah menjadi satuan wilayah yang lebih kecil atau menggunakan model yang memperhatikan aspek heterogenitas spasial

DAFTAR PUSTAKA

- Adetunji, A. B., Akande, O. N., Ajala, F. A., Oyewo, O., Akande, Y. F., & Oluwadara, G. (2022). House Price Prediction using Random Forest Machine Learning Technique. *Procedia Computer Science*, 199, 806–813. <https://doi.org/10.1016/j.procs.2022.01.100>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Christy, E., & Suryowati, K. (2021). Analisis Klasifikasi Status Bekerja Penduduk Daerah Istimewa Yogyakarta Menggunakan Metode Random Forest. *Jurnal Statistika Industri Dan Komputasi*, 6(1), 69–76.
- Díaz-Uriarte, R., & Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7, 1–13. <https://doi.org/10.1186/1471-2105-7-3>
- Direktorat Penilaian Tanah dan Ekonomi Pertanian. (2022). *Petunjuk Teknis Direktorat Penilaian Tanah dan Ekonomi Pertanian Tahun 2022*. Kementerian Agraria dan Tata Ruang/Badan Pertanahan Nasional.
- Genuer Robin, & Poggi Jean-Michel. (2020). *Random Forests with R*. Springer. <http://www.springer.com/series/6991>
- Gloude-mans, R. . (1982). *Simplified salesbased models for condominium/townhouse valuation*.
- Goldstein, B. A., Polley, E. C., & Briggs, F. B. S. (2011). Random forests for genetic association studies. *Statistical Applications in Genetics and Molecular Biology*, 10(1), 32. <https://doi.org/10.2202/1544-6115.1691>
- IAAO. (2013). *Standard on Ratio Studies*. International Association of Assessing Officers. <http://www.iaao.org>
- IAAO. (2017). *Standard on Mass Appraisal of Real Property*. International Association of Assessing Officers. www.iaao.org
- IAAO. (2018). *Standard on Automated Valuation Models (AVMs)*. International Association of Assessing Officers. www.iaao.org

- Insukindro. (1992). Pembentukan Model Dalam Penelitian Ekonomi. *Journal of Indonesian Economy and Business*, 7, 18. <https://journal.ugm.ac.id/jieb/article/view/40075>
- KEPI dan SPI. (2018). *Kode Etik Penilai Indonesia dan Standar Penilai Indonesia: Vol. Edisi VII (VII–2018)*. MAPPI.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Team, R. C., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., & Hunt, T. (2023). *Package ‘caret.’* <https://doi.org/10.32614/CRAN.package.caret>
- Montaño Moreno, J. J., Palmer Pol, A., Sesé Abad, A., & Cajal Blasco, B. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema*, 25(4), 500–506. <https://doi.org/10.7334/psicothema2013.23>
- Muningsih, E. (2022). Kombinasi Metode K-Means Dan Decision Tree Dengan Perbandingan Kriteria Dan Split Data. *Jurnal Teknoinfo*, 16(1), 113. <https://doi.org/10.33365/jti.v16i1.1561>
- Musianto, L. S. (2002). Perbedaan Pendekatan Kuantitatif Dengan Pendekatan Kualitatif Dalam Metode Penelitian. *Jurnal Manajemen Dan Wirausaha*, 4(2), 123–136. <https://doi.org/10.9744/jmk.4.2.pp.123-136>
- Nursalman, & Mustikasari. (2020). Pengaruh Penyetelan Hyperparameter Terhadap Kinerja Prediksi Random Forest pada Pendeteksian Spam. In *JURNAL INSTEK Informatika Sains dan Teknologi* (Vol. 5, Issue 2, pp. 149–158).
- Páez, A., Uchida, T., & Miyamoto, K. (2001). Spatial Association and Heterogeneity Issues in Land Price Models. *Urban Studies*, 38(9), 1493–1508. <http://www.jstor.org/stable/43196723>
- Pamungkas, M. R., Soedwihajono, S., & Miladan, N. (2019). Kesiapan Kawasan Commercial Strip Surakarta Utara Bagian Barat Sebagai Pusat Aktivitas Baru. *Desa-Kota*, 1(2), 167. <https://doi.org/10.20961/desa-kota.v1i2.14756.167-176>
- Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(3). <https://doi.org/10.1002/widm.1301>
- Putra, P. H., Azanuddin, A., Purba, B., & Dalimunthe, Y. A. (2023). Random forest and decision tree algorithms for car price prediction. *Jurnal Matematika Dan Ilmu Pengetahuan Alam LLDikti Wilayah 1 (JUMPA)*, 3(2), 81–89. <https://doi.org/10.54076/jumpa.v3i2.305>
- Ren, P., Li, Y., & You, K. (2021). Residents Demands for Urban Retail: Heterogeneity in Housing Structure Characteristics, Price Quantile, and Space. *Land*, 10(12). <https://doi.org/10.3390/land10121321>
- Song, J., Gao, Y., Yin, P., Li, Y., Li, Y., Zhang, J., Su, Q., Fu, X., & Pi, H. (2021). The random forest model has the best accuracy among the four pressure ulcer prediction models using machine learning algorithms. *Risk Management and Healthcare Policy*, 14, 1175–1187. <https://doi.org/10.2147/RMHP.S297838>
- Suliztia, M. L. (2020). Penerapan Analisis Random Forest Pada Prototype Sistem Prediksi Harga Kamera Bekas Menggunakan Flask. In *Fakultas Matematika dan Ilmu Pengetahuan Alam*. Universitas Islam Indonesia.
- Wooseok, K., Eunchan, K., & Wookjae, H. (2022). The Association Between SOC and Land Prices Considering Spatial Heterogeneity Based on Finite Mixture Modeling. *ArXiv*, 9, 53–78. <http://dx.doi.org/10.15793/kspr.2022.114..004>
- Wright, M. N., Dankowski, T., & Ziegler, A. (2017). Unbiased split variable selection for

random survival forests using maximally selected rank statistics. *Statistics in Medicine*, 36(8), 1272–1284. <https://doi.org/10.1002/sim.7212>

Wright, M. N., Wager, S., & Probst, P. (2023). *ranger: A Fast Implementation of Random Forests*. <https://doi.org/10.32614/CRAN.package.ranger>