



Penilaian Properti: Penggunaan *Machine learning* untuk Prediksi Nilai Sewa

Cindy Clara Afrisca^{1*}, Hanif Noer Rofiq², Darmawan Dwi Atmoko³,
Kementerian Keuangan, Republik Indonesia,

cindyclarisca@gmail.com¹, hanif.rofiq@kemenkeu.go.id², darmawan.atmoko@kemenkeu.go.id³

*penulis korespondensi

Keywords: *Machine learning, Random forest, property valuation, Rental value prediction, Real estate, Automated valuation model*

ABSTRACT

The rising demand for rental spaces, such as hotel lodging rooms, has led property owners in high-potential areas to begin renting out their properties to meet this need. The Directorate General of State Assets under the Ministry of Finance, which manages state-owned assets, oversees approximately 1,100 room units with potential for rental. In alignment on increasing non-tax state revenue from state-owned assets, offering lodging room rentals as a form of BMN utilization presents a promising opportunity for expansion. To facilitate efficient rental services for these BMN rooms, a quick and accurate valuation process is essential. However, establishing fair rental prices for both property owners and tenants poses a significant challenge. This study, therefore, aims to develop automated valuation model. The study leverages a diverse array of machine learning algorithms such as Ridge, Linear Regression, Random Forest, Lasso, Support Vector Regression, Elastic Net, and Extreme Gradient Boosting—to enhance the accuracy of rental price predictions by considering multiple independent variables. The research focuses on rental markets in Jakarta, Bogor, and Bandung, using a dataset of fewer than 500 rental room samples sourced from marketplaces. These data points are adjusted for state-owned asset variables, followed by feature engineering. The performance of the various algorithms is then compared. The results indicate that the best-performing model and the most important features vary by city, but overall, the Random Forest model delivers strong performance, with room size, road width, and room quality emerging as the most influential factors.

Kata Kunci: *Machine learning, random forest, penilaian properti, prediksi nilai sewa, real estate, automated valuation model*

ABSTRAK

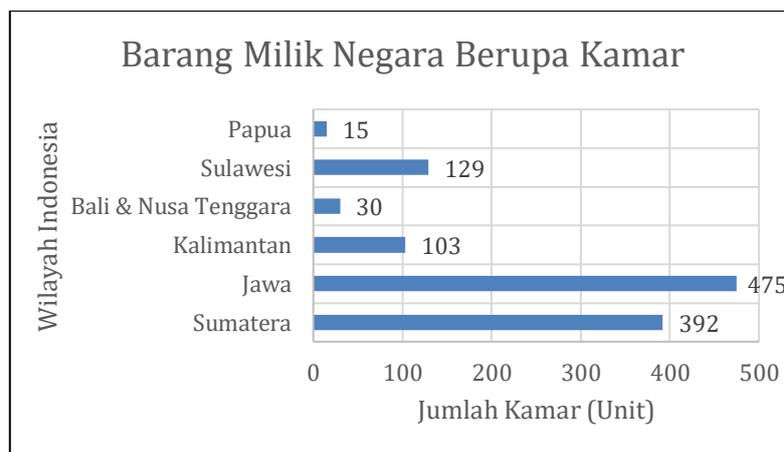
Bidang persewaan seperti kamar penginapan hotel semakin meningkat kebutuhannya. Pemilik properti di berbagai wilayah yang berpotensi mulai menyewakan properti mereka untuk memenuhi permintaan ini. Selanjutnya, Direktorat Jenderal Kekayaan Negara,

Kementerian Keuangan sebagai pengelola aset negara memiliki aset berupa kamar sekitar 1.100 unit yang berpotensi untuk disewakan. Maka dari itu, selaras dengan upaya meningkatkan penerimaan negara bukan pajak dari aset Barang Milik Negara, sehingga mengoptimalkan pelayanan pemanfaatan Barang Milik Negara berupa sewa kamar penginapan menjadi salah satu cara yang baik untuk di kembangkan. Demikian guna mendukung pelayanan sewa Barang Milik Negara kamar tersebut di perlukan proses penilaian yang cepat dan tetap akurat. Namun, menentukan nilai yang adil bagi pemilik properti dan penyewa adalah sebuah tantangan. Oleh karena itu, penelitian ini bertujuan membentuk *automated valuation model* untuk prediksi nilai sewa properti. Prediksi nilai sewa properti sensitif terhadap beberapa parameter bebas, karena itu penelitian ini akan menggunakan beberapa algoritma *machine learning* seperti *Ridge*, *Linear Regression*, *Random Forest*, *Lasso*, *Support Vector Regression*, *Elastic Net*, dan *Extreme Gradient Boosting*. Penelitian ini dilakukan pada ruang lingkup kota Jakarta, Bogor, dan Bandung, dengan jumlah data di bawah 500 dengan sampel data sewa kamar yang didapat dari marketplace yang telah disesuaikan variabelnya dengan kondisi barang milik negara, rekayasa fitur, dan kemudian membandingkan hasil dari setiap algoritma. Hasil yang didapat, model terbaik dan *feature importance* dari masing-masing kota berbeda namun secara keseluruhan model Random Forest memiliki hasil yang cukup memuaskan dengan ukuran kamar, lebar jalan, dan kualitas kamar menjadi variabel yang paling penting.

PENDAHULUAN

Indonesia mempunyai laju pertumbuhan penduduk yang sangat pesat. Tercatat bahwa Indonesia merupakan peringkat ke-4 sebagai Negara dengan populasi terbanyak di dunia (Roser, Ritchie, & Ortiz-Ospina, 2019). Kepadatan penduduk di setiap daerah dapat berbeda-beda, ini dikarenakan penduduk relatif mencari tempat yang lebih baik dalam perkumpulan ke suatu tempat yang memiliki banyak sumber daya dan fasilitas kehidupan (Suhardi, 2009). Provinsi dengan kepadatan penduduk tertinggi adalah kota-kota besar seperti DKI Jakarta, Surabaya, Bogor, Bandung, dan Yogyakarta. Kepadatan penduduk yang cukup tinggi berbanding lurus dengan peningkatan kebutuhan ruang pemukiman untuk kegiatan bisnis, hiburan atau pun kunjungan lainnya. Oleh karena itu, diperlukan suatu tempat hunian praktis untuk menekan jumlah lahan hunian ataupun cara untuk dapat memaksimalkan objek dan fasilitas yang telah ada.

Pada tahun 2023, diketahui bahwa pendapatan untuk pemanfaatan Barang Milik Negara (BMN) sekitar 708 miliar rupiah memberikan kontribusi PNBPN (Kementerian Keuangan, 2023). Sebagai contohnya, sewa barang milik negara berupa kamar termasuk kedalam jenis pendapatan tersebut. Berdasarkan Peraturan Menteri Keuangan Nomor 115 Tahun 2020 Tentang Pemanfaatan BMN, merujuk kepada itu Sewa Aset BMN merupakan salah satu bentuk pemanfaatan yang paling sering dimohonkan oleh Pengguna Barang kepada Pengelola Barang. Kemudian, dari hasil observasi yang telah dilakukan melalui survei diperoleh indikasi potensi BMN yang dapat di manfaatkan untuk sewa kamar penginapan. Berikut adalah gambaran prospek BMN yang dapat dimanfaatkan untuk sewa BMN.



Gambar 1. Barang Milik Negara berupa Kamar

Sumber: Direktorat Penilaian, 2022 (sumber: data sekunder diolah)

Data pada gambar 1, Pada tahun 2022 terdapat BMN berupa kamar penginapan dengan potensi tertinggi di Pulau Jawa (475), dan daerah lainnya yaitu Sumatera (392), Kalimantan (103), Bali dan Nusa Tenggara (30), Sulawesi (129), dan Papua (15) (Direktorat Penilaian, 2022). Demikian, menunjukkan bahwa masih terdapat potensi utilisasi atas BMN sewa yang dapat menjadi alternatif untuk meningkatkan penerimaan negara melalui setoran ke kas negara sebagai PNBPN dari pelayanan pemanfaatan sewa BMN. Dalam melakukan transaksi sewa dibutuhkan penilaian yang fit dan proper. Penilaian terhadap aset properti sangat diperlukan baik di kalangan swasta maupun pemerintah, dimana tujuannya untuk meningkatkan akuntabilitas dan terciptanya tertib administrasi dalam mendukung optimalisasi pengelolaan aset ke arah yang lebih baik dan modern. Melihat urgensi adanya potensi optimalisasi atas

barang milik negara berupa kamar maka penelitian ini dilakukan agar dapat menemukan nilai sewa barang milik negara berupa kamar yang kompetitif dengan pasar dan menempuh proses waktu pengerjaan yang efektif. Untuk mengimplmentasikan hal tersebut diperlukan sebuah alat data analitik sebagai serangkaian aplikasi yang menggunakan basis intelegensi otomatis. Data mining adalah salah satu proses ekstraksi informasi atau pola dari database dengan jumlah data yang besar telah memberikan keuntungan dalam berbagai penelitian (Chapman dkk, 2000). Secara teknis, *data mining* dapat digunakan untuk memprediksi nilai melalui teknik klasifikasi dan regresi untuk menemukan model yang menggambarkan dan membedakan pola atau kelas-kelas data (Liaw & Wiener, 2002; Susanto & Sudiyanto, 2014). Pengelompokan teknik regresi dikategorikan kedalam beberapa algoritma diantaranya *ridge*, *linear regression*, dan *random forest (rf)*. Dalam hal ini, *data mining* berfungsi sebagai *machine learning* yang berjalan di atas *database*, optimisasi *query*, karena dapat mengakses data yang ukurannya besar (Alpaydin, 2020). Salah satu kemampuan *data mining* yang dapat dirasakan pada proses penilaian sewa yaitu terutama pada proses mengekstraksi pengetahuan yang diinginkan dari sebuah basis data, menganalisis, memahami, atau bahkan menggambarkan melalui pencarian pengetahuan dalam basis data dalam proses identifikasi pola-pola yang tepat serta berpotensi manfaat, sehingga dapat dipahami secara mudah dan efektivitas waktu penilaian (Direktorat Jenderal Kekayaan Negara, 2021).

Prediksi nilai sewa telah menjadi salah satu permasalahan dalam percepatan layanan pengelolaan BMN karena penilaian sewa BMN konvensional tergantung pada biaya dan perbandingan harga pasar pada beberapa kasus tidak memenuhi standar yang diterima dan proses sertifikasi. Prediksi nilai sewa BMN kamar yang akurat sangat penting karena secara langsung memengaruhi Pendapatan Negara Bukan Pajak (PNBP) dari pengelolaan aset BMN. Sebab itu, diperlukan identifikasi faktor-faktor yang berpengaruh signifikan dalam menentukan nilai sewa BMN kamar untuk memastikan pengelolaan yang optimal dan mendukung peningkatan pendapatan negara. Sewa kamar dipengaruhi oleh berbagai faktor, seperti lokasi, ukuran kamar, kualitas, dan fasilitas yang tersedia (Ho dkk., 2020; Sangha, 2021). Oleh karena itu, pemutakhiran penilaian sewa kamar memerlukan pembaruan metodologi, model, dan teknologi untuk menjamin akurasi penilaian. Salah satu pendekatan yang dapat digunakan adalah penerapan proses data mining dengan dukungan algoritma machine learning, yang mampu menganalisis data dalam jumlah besar dan menemukan pola atau hubungan yang relevan secara efektif (Chapman dkk., 2000; Alpaydin, 2020).

Machine learning, sebagai cabang *artificial intelligence*, digunakan untuk memprediksi nilai sewa dengan algoritma seperti *random forest*, *ridge regression*, dan *linear regression* (Alpaydin, 2020; Kim, 2019; Liaw & Wiener, 2002). *Framework* seperti *Cross-Industry Standard Process for Data Mining* (CRISP-DM) dirancang untuk mendukung implementasi *machine learning* yang terstruktur dalam proses penilaian (Chapman dkk., 2000). Meskipun metode konvensional telah memenuhi standar, *machine learning* menawarkan solusi otomatisasi yang meningkatkan efisiensi dan akurasi penilaian (Sangha, 2021; Ho dkk., 2020). Salah satu keunggulannya adalah mendukung keadilan transaksi dengan penyesuaian berbasis *data-driven decision making*, yang mempercepat proses penilaian (Direktorat Jenderal Kekayaan Negara, 2021; Chapman dkk., 2000). Selain itu, *machine learning* mendukung pemutakhiran penilaian melalui permodelan regresi otomatis seperti linear dan *ridge regression*, memastikan metodologi yang relevan dan akurat (Kim, 2019; Dewi, 2011).

Penelitian ini bertujuan untuk mengembangkan model prediksi yang dapat digunakan untuk memprediksi nilai sewa BMN kamar. Evaluasi model dilakukan dengan mengukur nilai *Root Mean Square Error* (RMSE), *Mean Absolute Error* (MAE), *RMSE cross-validation*, dan *R² score (coefficient of determination)*. Selain itu, penelitian ini juga mengidentifikasi faktor-faktor penting yang memengaruhi nilai sewa kamar menggunakan algoritma regresi seperti *Random Forest* dan *Ridge Regression*. Data yang digunakan dalam penelitian ini dikumpulkan dari *marketplace* di wilayah Pulau Jawa, dengan fokus pada kota Jakarta, Bogor, dan Bandung sebagai wilayah dengan potensi sewa terbesar, untuk mendukung prediksi nilai sewa BMN kamar.

Model prediksi ini menggunakan fitur intrinsik properti seperti ukuran kamar, kondisi interior, jenis tempat tidur, dan fasilitas (amenities), serta variabel ekstrinsik seperti lokasi, jumlah pengunjung, dan kualitas. Tahapan penelitian mencakup pra-pemrosesan data dan pemeriksaan kualitas menggunakan statistik deskriptif. Proyek regresi dilakukan menggunakan *Jupyter Notebook* dengan pustaka *Python* terkait, termasuk *Pandas*, *scikit-learn* (sklearn), *Ridge*, *Linear Regression*, *Random Forest* (RF), *Lasso*, *Support Vector Regression* (SVR), *Elastic Net*, dan *Extreme Gradient Boosting* (XGBoost).

Sebelum penelitian ini, telah terdapat beberapa penelitian serupa yang menerapkan algoritma *machine learning* dalam penilaian properti. Penelitian tersebut menggunakan berbagai jenis data dan metode prediksi sesuai dengan tujuan penelitian. Metode *Random Forest* diketahui dapat digunakan dalam *driver analysis*, yaitu analisis untuk mengetahui pengaruh variabel bebas terhadap variabel terikat. Hasil penelitian menunjukkan bahwa berdasarkan nilai *Mean Decrease Gini* (MDG), stabilitas model *Random Forest* dapat dicapai dengan ukuran lebih dari 500 *decision tree*. Selain itu, berdasarkan rata-rata MDG dari 1.000 iterasi *Random Forest*, analisis tetap stabil meskipun menggunakan berbagai ukuran jumlah variabel bebas (Dewi, 2011).

Penelitian lainnya menunjukkan bahwa *Random Forest* memiliki akurasi model yang lebih baik dibandingkan dengan *Regularized Linear Regression* (Gong dkk., 2018). Kombinasi *Stepwise* dan *Support Vector Machine* diusulkan sebagai algoritma prediksi paling akurat untuk memprediksi harga rumah di Melbourne, dengan faktor penting seperti jumlah kamar tidur, jarak ke CBD, koordinat, dan jenis rumah (Phan, 2018). Metode *Random Forest* biasa dibandingkan dengan *Guided Regularized Random Forest* (GRRF) menggunakan data *geodatabase* penginderaan jauh, dan hasilnya menunjukkan bahwa GRRF hanya meningkatkan akurasi sebesar 2% dibandingkan *Random Forest* biasa, sehingga perbedaannya hampir tidak signifikan (Izquierdo-Verdiguier & Zurita-Milla, 2020). Model *tree* dengan *boosting* terbukti memberikan nilai MAE terendah dan *R² score* tertinggi, menjadikannya model prediksi paling sukses untuk penilaian *real estate* dengan jumlah variabel yang sedikit (Sangha, 2021).

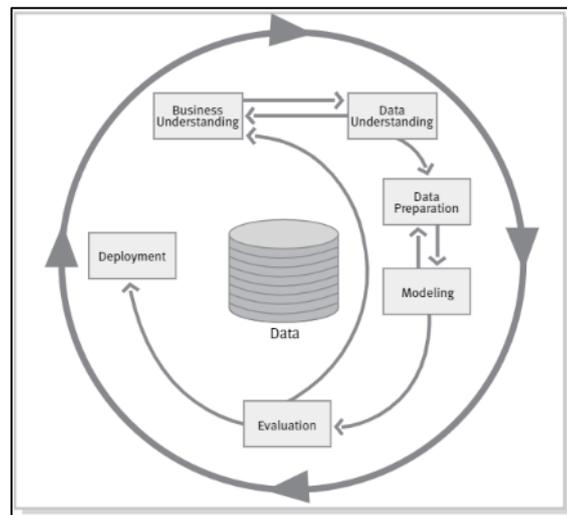
Penulis mengkaji literatur yang relevan tersebut tentang *automated valuation models* untuk mencapai akurasi prediksi yang sebanding, meskipun terdapat keterbatasan jumlah dan konteks data dalam studi ini. Hasil penelitian menunjukkan bahwa model terbaik dengan algoritma *Random Forest* untuk estimasi nilai sewa kamar memiliki *R² score* sebesar 0,49 dengan *NRMSE cross-validation* sebesar 0,13.

METODE PENELITIAN

Cross-industry standard process for data mining (CRISP-DM)

Penelitian ini menerapkan metodologi *Cross-Industry Standard Process for Data Mining* (CRISP-DM) sebagai pendekatan utama untuk memprediksi nilai sewa kamar. Kerangka kerja ini

menawarkan langkah-langkah yang terstruktur, fleksibel, dan terbukti efektif dalam berbagai proyek data mining, termasuk penilaian properti. Melalui enam tahapan utama yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Metode ini mendukung pengembangan model secara iteratif, sehingga meningkatkan akurasi dan relevansi hasil (Chapman dkk., 2000). Metodologi ini juga memfasilitasi identifikasi faktor penting, persiapan data, dan evaluasi model dengan pendekatan yang terukur, menjadikannya solusi ideal untuk penelitian data yang kompleks. Proses metodologi ini terdiri dari 6 tahapan sebagai berikut.



Gambar 2. *Cross-Industry Standard Process* untuk *Data mining* (Chapman, 2000)

Business Understanding

Pada tahap *business understanding*, beberapa langkah yang dilakukan meliputi memahami kebutuhan serta tujuan dari sudut pandang proses bisnis pengelolaan BMN di DJKN, kemudian menerjemahkan pengetahuan tersebut ke dalam bentuk pendefinisian masalah pada data mining. Selanjutnya, rencana dan strategi disusun untuk mencapai tujuan data mining secara efektif. Berdasarkan Peraturan Direktur Jenderal Kekayaan Negara Nomor 4 Tahun 2021 tentang Penilaian Sewa, proses penilaian properti salah satunya menggunakan pendekatan data pasar, yang melibatkan teknik penyesuaian nilai melalui analisis regresi sebagai bagian dari metode statistik. Maka, berdasarkan ketentuan yang ada, penelitian ini menggunakan pendekatan data pasar untuk memprediksi nilai sewa dengan penyesuaian melalui regresi statistik yang diterapkan menggunakan model berbasis *machine learning*.

Data Understanding

Tahapan ini mencakup pengumpulan, deskripsi, dan evaluasi kualitas data. Data yang digunakan berasal dari *marketplace*, mencakup variabel properti intrinsik dan ekstrinsik pada periode tertentu. Variabel bebas meliputi data koordinat, jarak ke CBD, lebar jalan, jumlah pesaing, jumlah tamu, kualitas kamar, interior, ukuran kamar, serta fasilitas seperti wifi, kamar mandi, amenities, sarapan, furnitur, kulkas, televisi, ketel listrik, jenis tempat tidur, dan pengatur suhu ruangan. Variabel dependen dalam penelitian ini adalah harga sewa. Klasifikasi variabel yang relevan dilakukan untuk memastikan kesesuaiannya dengan kondisi BMN berupa kamar. Proses verifikasi klasifikasi didukung oleh survei lapangan pada BMN berupa kamar di wilayah Bogor.

Wilayah Jakarta, Bogor, dan Bandung dipilih sebagai sampel penelitian karena ketiganya merupakan pusat kegiatan ekonomi, sosial, dan wisata di Pulau Jawa, yang memiliki tingkat permintaan sewa kamar yang tinggi dan beragam. Jakarta sebagai ibu kota negara memiliki dinamika pasar properti yang kompleks, sementara Bogor dan Bandung merupakan kota penyangga utama dengan karakteristik pasar yang melengkapi Jakarta, seperti tingginya potensi wisatawan dan penyewa jangka pendek. Ketiga wilayah ini memberikan representasi yang baik untuk menganalisis variasi nilai sewa kamar di kawasan perkotaan dan sub-urban. Data yang digunakan mencakup pembaruan terakhir pada 15 Juni 2022, dengan jumlah data masing-masing kota adalah Jakarta sebanyak 375, Bogor sebanyak 194, dan Bandung sebanyak 439, serta total 17 variabel untuk setiap wilayah.

Data Preparation

Dalam tahapan ini dilakukan pembangunan *dataset* akhir dari berupa data mentah menjadi data yang siap olah. Hal yang pertama dilakukan adalah melakukan *Exploratory Data Analytics* untuk melihat dengan jelas bagaimana data dimiliki.

Tabel 1. *Data Observation* (sumber: data sekunder diolah)

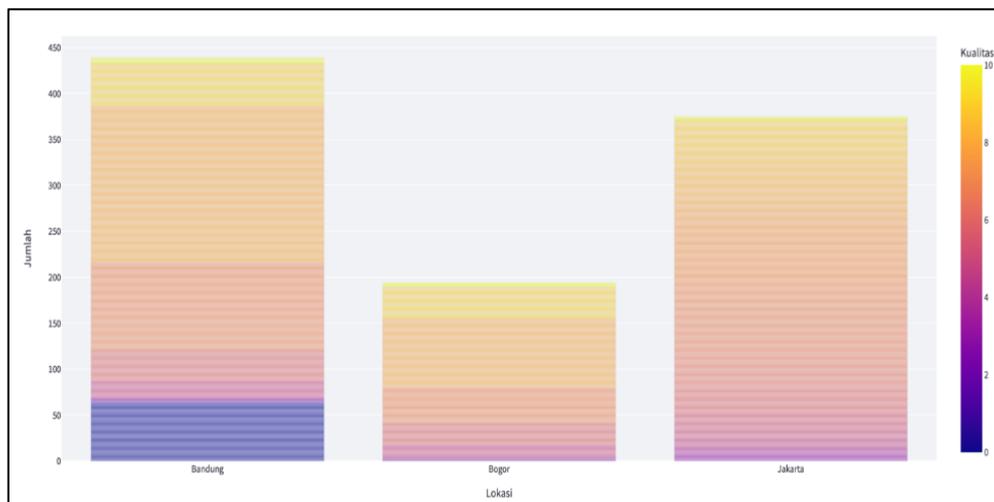
Variabel	Tipe Data	Role
Harga	Integer	<i>Target/Variabel terikat</i>
Lebar_Jalan	Float	<i>Feature/Variabel bebas</i>
Pesaing	Integer	<i>Feature/Variabel bebas</i>
Kualitas	Float	<i>Feature/Variabel bebas</i>
Guest	Integer	<i>Feature/Variabel bebas</i>
Interior	Integer	<i>Feature/Variabel bebas</i>
Ukuran_Kamar	Integer	<i>Feature/Variabel bebas</i>
Wifi	Integer	<i>Feature/Variabel bebas</i>
Bathroom	Integer	<i>Feature/Variabel bebas</i>
Toiletries	Integer	<i>Feature/Variabel bebas</i>
Breakfast	Integer	<i>Feature/Variabel bebas</i>
Furniture	Integer	<i>Feature/Variabel bebas</i>
Kulkas	Integer	<i>Feature/Variabel bebas</i>
Televisi	Integer	<i>Feature/Variabel bebas</i>
Ketel_Listrik	Integer	<i>Feature/Variabel bebas</i>
Jenis_Tempat_Tidur	Integer	<i>Feature/Variabel bebas</i>
Air Conditioning (AC)	Integer	<i>Feature/Variabel bebas</i>

Variabel-variabel penelitian dapat dijelaskan sebagai berikut: Harga merupakan nilai sewa kamar per hari. Lebar jalan menunjukkan ukuran jalan utama di depan lokasi unit yang disewakan. Kategori pesaing mengindikasikan tingkat persaingan sewa kamar di suatu kota, yang diklasifikasikan menjadi tinggi (lebih dari 10 tempat penginapan), sedang (5–10 penginapan), dan rendah (kurang dari 5 penginapan). Kualitas kamar dinilai berdasarkan lima

aspek: kondisi fisik bangunan, struktur bangunan, kebersihan, keamanan, dan aksesibilitas. Kualitas dikategorikan sebagai baik (memenuhi semua aspek), sedang (memenuhi minimal 4 aspek), atau buruk (memenuhi minimal 3 aspek).

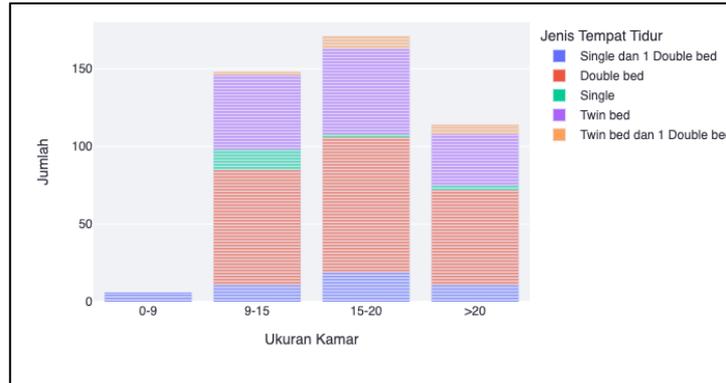
Jumlah pengunjung (*guest*) mengacu pada jumlah tamu dalam satu kamar. Interior dinilai berdasarkan estetika fisik kamar, dengan kategori baik, sedang, atau kurang baik. Ukuran kamar dinyatakan dalam meter persegi, sementara ketersediaan fasilitas seperti internet/wifi, kamar mandi dalam (*bathroom*), alat mandi (*toiletries*), sarapan (*breakfast*), furnitur (meja, kursi, dan lemari), kulkas, televisi, ketel air, dan pendingin ruangan (air conditioning/AC) dicatat sebagai variabel biner ada/tidak ada. Jenis tempat tidur diklasifikasikan sebagai *single bed*, *double bed*, *twin bed*, kombinasi *single* dan *double bed*, atau kombinasi *twin bed* dan *double bed*.

Berdasarkan informasi pada Tabel 2, data yang dikumpulkan ditransformasikan sesuai kategorinya menjadi variabel biner atau nominal menggunakan *machine learning*, sehingga tidak memerlukan pemetaan manual dari data kualitatif menjadi kuantitatif. Selanjutnya, distribusi data berdasarkan kualitas kamar dianalisis sebagaimana ditampilkan pada Gambar 3.



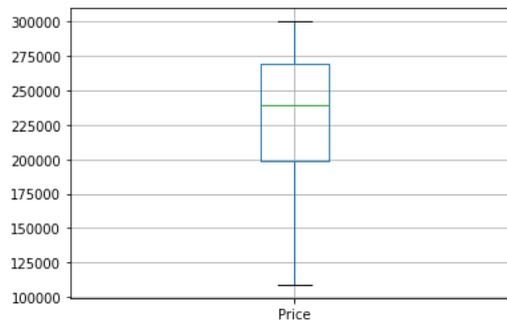
Gambar 3. Distribusi data berdasarkan kualitas (sumber: data sekunder diolah)

Berdasarkan Gambar 3, distribusi data berdasarkan kualitas kamar di wilayah Bogor dan Jakarta kurang baik karena memiliki sebaran kualitas kamar yang kurang variatif dibandingkan dengan Bandung, yang menunjukkan distribusi data yang lebih baik. Namun, jika dianalisis lebih mendalam dengan membandingkan data ukuran kamar dan jumlah tempat tidur di Bandung, distribusi jenis data yang dikumpulkan masih tergolong tidak seimbang (*imbalanced*), karena hanya mencakup 5 dari 7 jenis tempat tidur yang tersedia.



Gambar 4. Jumlah Ukuran Kamar dan Jenis Tempat Tidur (sumber: data sekunder diolah)

Tahapan selanjutnya adalah *data cleaning*. Pada proses *data gathering*, seleksi awal telah dilakukan terhadap data yang akan digunakan sebagai *training data*, termasuk jenis penginapan, kualitas, dan rentang harga. Oleh karena itu, data yang telah dikumpulkan relatif bersih dari *outlier*, sebagaimana ditunjukkan pada Gambar 5. Namun, data tersebut masih mengandung *missing values*. Untuk menangani *missing values*, langkah yang dilakukan adalah dengan mengisi nilai yang hilang menggunakan metode imputasi yang sesuai dengan karakteristik data.

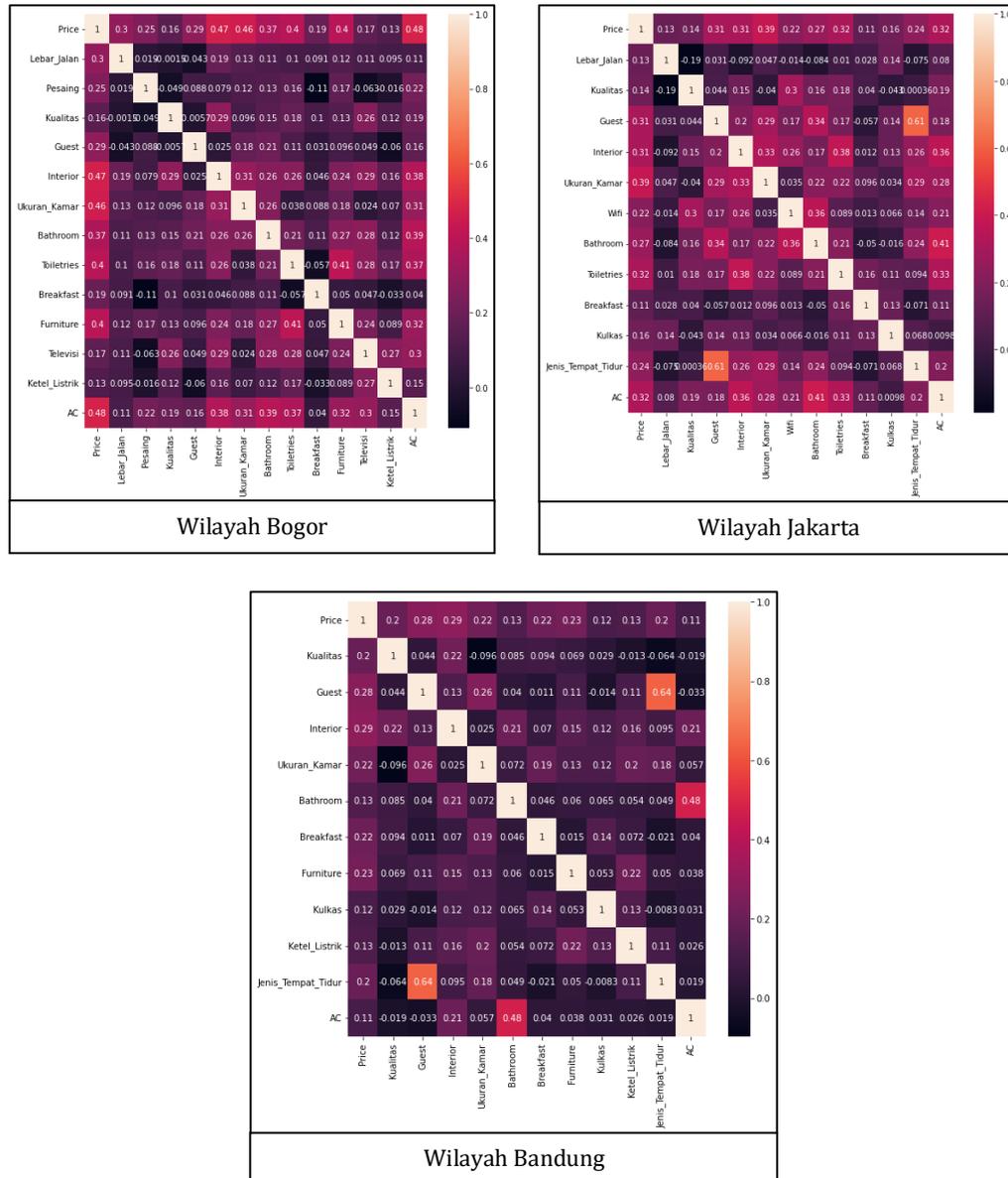


Gambar 5. *Boxplot* harga sewa kamar dari data wilayah Jakarta (sumber: data sekunder diolah)

Observasi korelasi antara variabel *Price* dan variabel bebas dilakukan menggunakan *heatmap* pada Gambar 6. Hasil analisis menunjukkan bahwa korelasi antara variabel terikat dan variabel bebas bervariasi di setiap kota. Secara umum, variabel *Interior* memiliki korelasi tinggi dengan *Price* di semua kota, menunjukkan pentingnya estetika kamar dalam menentukan harga sewa. Korelasi tertinggi di masing-masing kota adalah penyejuk ruangan atau AC di Bogor (0,47), ukuran kamar di Jakarta (0,39), dan interior di Bandung (0,29).

Selain itu, variabel lain seperti *Kualitas*, *Guest*, dan fasilitas (*toiletries* dan *Bathroom*) juga menunjukkan korelasi positif terhadap *Price*, meskipun tidak sekuat variabel utama. Hal ini mengindikasikan bahwa faktor-faktor tersebut tetap berkontribusi pada nilai sewa kamar.

Setelah analisis, data dibagi dengan rasio 80:20, di mana 80% digunakan sebagai *training data* untuk melatih model prediksi, dan 20% sisanya sebagai *testing data* untuk mengevaluasi akurasi model. Pembagian ini memastikan model dapat mempelajari pola dengan baik dan memprediksi data baru secara optimal.



Gambar 6. Map Correlation (sumber: data sekunder diolah)

Modeling

Pada bagian ini, algoritma machine learning digunakan untuk membangun sistem rekomendasi yang menyarankan nilai wajar sewa kamar berdasarkan informasi properti yang tersedia di pasar. Sistem ini juga memungkinkan penyesuaian nilai sewa berdasarkan hasil prediksi sesuai kebutuhan. Dalam kasus ini, model regresi yang diterapkan meliputi *Ridge Regression*, *Linear Regression*, *Random Forest (RF)*, *Lasso*, *Support Vector Regression (SVR)*, *Elastic Net*, dan *Extreme Gradient Boosting (XGBoost)*. Dengan pembahasan 4 model regresi yang utama yaitu sebagai berikut:

a. Regresi Linear

Regresi linear digunakan untuk memperkirakan nilai kontinu, seperti biaya rumah, jumlah panggilan, total penjualan, dan variabel serupa. Metode ini bekerja dengan memasang garis terbaik yang menciptakan hubungan antara variabel bebas dan

variabel dependen. Garis terbaik ini, yang disebut sebagai regresi, direpresentasikan oleh sebuah persamaan linear $Y = a * X + b$. Dalam persamaan ini:

Y – Variabel terikat

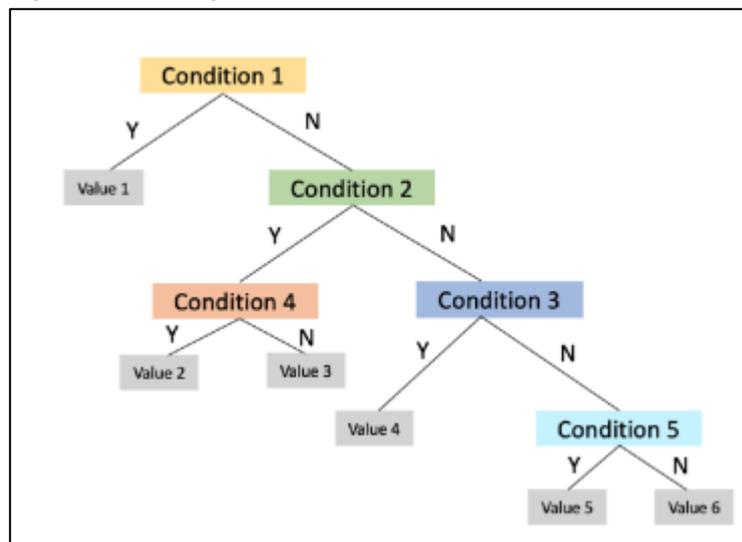
a – Kemiringan

X – Variabel Bebas

b – Intersep

Koefisien a dan b dalam regresi linear dihitung dengan meminimalkan jumlah kuadrat jarak antara garis regresi dan titik data (Alpaydin, 2020). Regresi linear memiliki berbagai bentuk, termasuk regresi linear sederhana, regresi linear berganda, dan variasi lainnya seperti ridge regression, yang mengatasi multikolinearitas dalam data (Kim, 2019). Ketika hubungan antara variabel bebas dan variabel dependen tidak linier, regresi polinomial dapat digunakan untuk menentukan kurva terbaik.

- b. *Ridge regression* mengintegrasikan regresi linear dengan elemen regularisasi berbasis L2 untuk mengurangi risiko *overfitting* parameter. Secara matematis, metode ini mencari nilai optimal yang meminimalkan jumlah kuadrat kesalahan residual ditambah dengan komponen regularisasi (Kim, 2019).
- c. *Random Forest* (RF) membentuk sejumlah *decision tree* selama proses *data train*, dan prediksi akhir diperoleh dengan mengombinasikan hasil prediksi dari masing-masing *tree*. Untuk mengurangi volatilitas yang sering terjadi pada algoritma dengan tingkat variasi tinggi, menggunakan metode pengambilan sampel acak dengan penggantian, yang dikenal sebagai *bagging* dalam *machine learning*. Algoritma ini biasanya memanfaatkan *decision tree*, di mana kumpulan data pelatihan terdiri dari fitur X dan keluaran Y, sementara *bagging* secara berulang memilih sampel acak dari kumpulan data pelatihan. sebanyak β kali $\beta \in \frac{1}{4}, 2, \dots, \beta P$ dan membangun pohon untuk sampel-sampel ini (Ho dkk., 2020).



Gambar 7. Algoritma *Decision Tree* (Ho dkk., 2020)

Dalam regresi, hasil akhir ditentukan dengan menghitung rata-rata prediksi yang dihasilkan oleh seluruh *decision tree*. *Random forest* memiliki sifat khas, seperti memilih secara acak sejumlah n sampel *bootstrap* dari data pengamatan yang tersedia daripada menggunakan seluruh *dataset*. Selain itu, hanya sebagian fitur yang dipilih

secara acak untuk setiap node, yang berfungsi untuk mengurangi kemungkinan overfitting (Liaw & Wiener, 2002).

- d. *Extreme Gradient Boosting* (XGBoost) merupakan salah satu algoritma pembelajaran mesin yang sering diterapkan dalam berbagai masalah dengan menawarkan kinerja yang sangat baik dalam banyak skenario (Chen & Guestrin, 2016). Algoritma ini sangat cocok untuk dataset berukuran besar, dengan waktu implementasi yang relatif singkat namun tetap mempertahankan tingkat akurasi yang tinggi. XGBoost bekerja secara iteratif dengan memanfaatkan *weak learner* untuk membentuk *strong learner*, di mana setiap model bergantung pada kesalahan prediksi model sebelumnya. Teknik ini menyesuaikan *tree* XGBoost dengan residual dan menerapkan metode pemisahan untuk mengevaluasi kualitas kinerja klaster, sekaligus menentukan ambang batas optimal bagi setiap fitur (Chen & Guestrin, 2016).

Evaluation

Setelah proses pemodelan selesai, hasil evaluasi dilakukan menggunakan metrik RMSE (*Root Mean Square Error*) dan nilai R² (*coefficient of determination*). RMSE dapat dinormalisasi untuk mempermudah perbandingan antara dataset atau model yang memiliki ukuran berbeda. Meskipun belum ada metode normalisasi yang disepakati secara universal dalam literatur, dua pendekatan yang sering digunakan adalah membagi RMSE dengan rata-rata data atau menggunakan rentang data, yang dihitung sebagai selisih antara nilai maksimum dan minimum observasi (Alpaydin, 2020). RMSE yang lebih rendah menunjukkan bahwa model memiliki varians sisa yang lebih kecil, yang mengindikasikan kinerja prediksi yang lebih baik. Dalam beberapa kasus, normalisasi ini dinyatakan dalam bentuk persentase dan dikenal sebagai *Normalized Root Mean Square Deviation* (NRMSD atau NRMSE) (Chen & Guestrin, 2016). Namun, rentang sampel dapat memengaruhi hasil normalisasi, terutama untuk dataset kecil, sehingga dapat menyulitkan proses perbandingan antar model (Chapman dkk., 2000). Menurut rumus di bawah ini:

$$\text{NRMSD} = \frac{\text{RMSD}}{y_{\max} - y_{\min}} \text{ or } \text{NRMSD} = \frac{\text{RMSD}}{\bar{y}}$$

Keterangan:

- NRMSD: *Normalized Root Mean Square Deviation* (RMSE yang telah dinormalisasi untuk memungkinkan perbandingan antara dataset atau model dengan skala yang berbeda)
- RMSD: *Root Mean Square Deviation* (metrik yang mengukur rata-rata kesalahan kuadrat antara nilai yang diprediksi dan nilai aktual dalam dataset)
- y_{\max} : Nilai maksimum dalam data observasi (nilai actual)
- y_{\min} : Nilai minimum dalam data observasi (nilai actual)
- \bar{y} : Rata-rata nilai aktual dari data observasi

Deployment

Tahapan ini dilakukan dengan visualisasi hasil permodelan, salah satunya bisa dengan pembuatan sistem aplikasi untuk perhitungan nilai sewa berdasarkan dari hasil permodelan.

HASIL DAN PEMBAHASAN

Perbandingan Algoritma Regresi

Dalam bagian ini, hasil evaluasi dari berbagai model yang digunakan akan dibandingkan. Hasil akhir ditunjukkan pada tabel di bawah ini.

Tabel 2. Perbandingan Algoritma *Machine learning* (sumber: data sekunder diolah)

Model	Bandung		Bogor		Jakarta	
	R2 Score	RMSE	R2 Score	RMSE	R2 Score	RMSE
<i>Linear Regression</i>	0.472	51956.356	0.026	47901.078	0.193	37980.841
<i>Ridge</i>	0.472	51966.787	0.049	47332.269	0.195	37912.079
<i>Lasso</i>	0.472	51956.732	0.026	47896.988	0.193	37977.390
<i>Elastic Net</i>	0.407	55065.657	0.126	45352.275	0.156	38839.124
<i>SVR</i>	0.483	51419.886	0.074	46706.145	0.055	41091.675
<i>Random Forest</i>	0.492	50960.780	0.323	39915.032	0.367	33616.473
<i>XGBoost</i>	0.479	51613.110	0.389	37933.895	0.287	35683.864

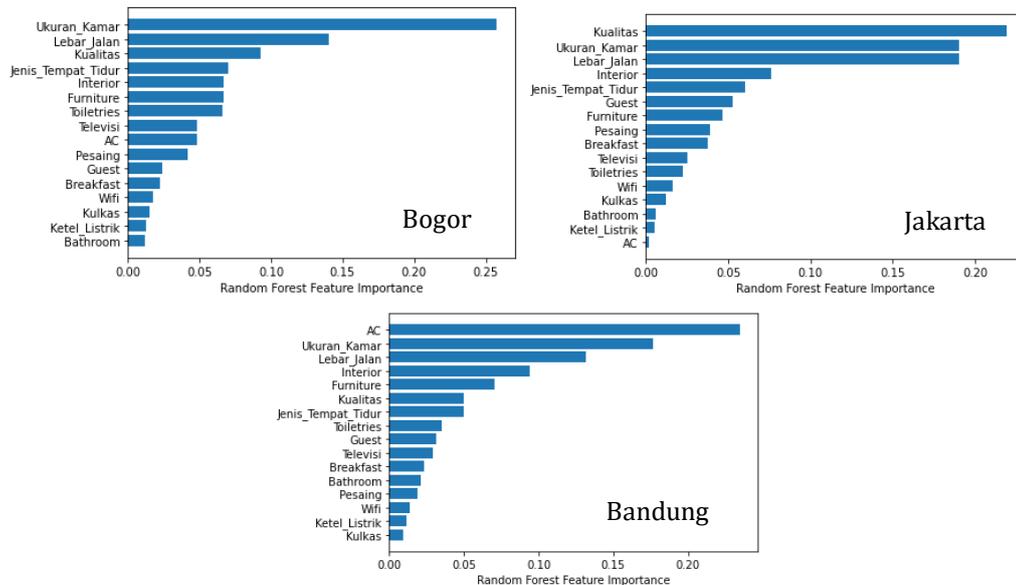
Dengan keterbatasan data yang tersedia, hasil *machine learning* menunjukkan bahwa model *Random Forest* memberikan performa terbaik dibandingkan model lain di Bandung ($R^2 = 0,49$), Bogor ($R^2 = 0,38$), dan Jakarta ($R^2 = 0,36$). Hal ini terlihat dari nilai R^2 yang lebih tinggi dan RMSE yang lebih rendah dibandingkan model lainnya. Untuk mengevaluasi seberapa baik nilai RMSE tersebut jika dibandingkan dengan distribusi harga (*Price distribution*) di masing-masing kota, dilakukan normalisasi terhadap nilai RMSE. Hasil nilai RMSE yang telah dinormalisasi dapat dilihat pada Tabel 3.

Tabel 3. Hasil Normalisasi RMSE (sumber: data sekunder diolah)

Kota	RMSE	NRMSE
Jakarta	33379,9	0,17
Bogor	39709,39	0,19
Bandung	51966,79	0,13

Berdasarkan Tabel 3, tanpa normalisasi, RMSE Bandung merupakan yang terbesar dibandingkan Bogor dan Jakarta. Namun, setelah normalisasi, RMSE Bandung menjadi yang terendah (0,13) karena rentang data di Bandung lebih luas dibandingkan kedua kota lainnya. Analisis selanjutnya membandingkan *Feature Importance* di setiap kota menggunakan model *Random Forest*. Meskipun *Random Forest* bukan model terbaik di Bogor, model ini dipilih karena jumlah data antar kota yang berbeda, serta perbedaan nilai R^2 dan RMSE yang tidak signifikan. Penggunaan model yang sama memberikan gambaran yang lebih konsisten mengenai fitur-fitur yang paling berkontribusi dalam prediksi harga sewa.

Feature Importance membantu menginterpretasikan model dengan menghitung kontribusi masing-masing fitur, sehingga dapat diidentifikasi variabel yang memiliki pengaruh terbesar dalam menentukan harga sewa di setiap kota.



Gambar 8. *Feature Importance* masing-masing kota (sumber: data sekunder diolah)
 Berdasarkan analisis *Feature Importance* dari model *Random Forest*, setiap kota memiliki variabel dominan yang berbeda dalam menentukan harga (*Price*). Di Bogor, variabel *Ukuran Kamar* memiliki pengaruh terbesar, diikuti oleh *Lebar Jalan* dan *Kualitas*. Sementara itu, di Jakarta, variabel *Kualitas* menjadi faktor paling penting, disusul oleh *Ukuran Kamar* dan *Lebar Jalan*. Berbeda dengan kedua kota lainnya, Bandung menunjukkan bahwa *AC* memiliki kontribusi terbesar dalam model, dengan *Ukuran Kamar* dan *Lebar Jalan* sebagai faktor berikutnya. Fasilitas tambahan seperti *Kulkas* dan *Bathroom* memiliki pengaruh kecil di semua kota. Perbedaan ini mencerminkan karakteristik unik dari pasar properti di masing-masing wilayah, di mana faktor tertentu lebih dominan dibandingkan yang lain. Namun, secara umum dapat disimpulkan kesamaan untuk *Ukuran Kamar* dan *Lebar jalan* yang menjadi variabel paling penting dalam menentukan nilai.

Supervisi Permodelan dalam Perhitungan Nilai Sewa

Permodelan yang telah dipilih yaitu model Bandung kemudian dilakukan deployment menjadi sebuah alat aplikasi perhitungan nilai sewa kamar. Guna melakukan supervisi atas model yang dibentuk maka Peneliti melakukan survei pasar melalui *marketplace* dan menganalisis sample harga sewa kamar yang paling terdekat dengan salah satu objek BMN yang berpotensi untuk disewakan, dengan hasil data perbandingan sebagai berikut:

Tabel 4. *Sample Data Sewa Kamar dan Objek BMN di Wilayah Bandung* (sumber: data sekunder diolah)

No	Keterangan	Data I	Data II	Data III	Objek BMN
1	Alamat/Lokasi	Jalan Asia Afrika No.34, 40111 Bandung, Indonesia	Jalan Asia Afrika No.34, 40111 Bandung, Indonesia	Jalan Dalem Kaum No. 130A , Asia Afrika, Bandung	Jl. Asia Afrika No.114 Kota Bandung

No	Keterangan	Data I	Data II	Data III	Objek BMN
	Sumber:	https://www.booking.com/hotel/id/zodiak-asiaafrika.id.html?aid=	https://www.booking.com/hotel/id/zodiak-asiaafrika.id.html?aid=	Data tabulasi, model sewa Kamar	Data BMN
2	Waktu Transaksi Sewa	2022	2022	2022	2022
3	Jenis Transaksi	Transaksi Sewa	Transaksi Sewa	Transaksi Sewa	Penilaian (Sewa)
4	Jarak ke CBD (satuan Meter)	257	257	392	428.36
5	lebar jalan depan	9	9	3	10
6	Ukuran	18	18	18	21
7	Guest	2	2	2	2
8	Wifi	Ada	Ada	Ada	Tidak Ada
9	Bathroom	Ada	Ada	Ada	Ada
10	Jenis Bed	Double	Double or twin	Double or twin	Double or twin
11	Toiletries	Ada	Ada	Ada	Tidak Ada
12	Sarapan	Ada	Ada	Tidak Ada	Tidak Ada
13	Furniture	Ada	Ada	Ada	Ada
14	Kulkas	Tidak Ada	Tidak Ada	Tidak Ada	Tidak Ada
15	TV Cable	Ada	Ada	Ada	Tidak Ada
16	TV	Ada	Ada	Ada	Tidak Ada
17	Ketel	Tidak Ada	Tidak Ada	Tidak Ada	Tidak Ada
18	AC	Ada	Ada	Ada	Ada
19	Interior	Interior Sangat Baik	Interior Sangat Baik	Interior Baik	Interior Baik
20	Pesaing	Tinggi	Tinggi	Sedang	Tinggi
21	Kualitas	Baik	Baik	Baik	Baik
	Harga sewa	Rp265.000	Rp265.000	Rp210.000	-

Pada tabel 4 rata-rata harga sewa harian kamar berkisar Rp246.000 untuk spesifikasi dan lokasi yang serupa dengan objek BMN. Implementasi hasil model ke dalam aplikasi dengan tampilan yang terbagi menjadi 2 bagian yaitu variabel eksternal dan variabel properti dengan *field input* yang berupa kuantitatif dan kualitatif mengikuti jenis data yang dipergunakan dalam melakukan kalkulasi harga sewa objek BMN berupa kamar seperti pada gambar 9. Hasil dengan model algoritma *Random Forest* untuk estimasi nilai sewa objek BMN berupa kamar sebesar Rp242.256,72 Dengan demikian, model dapat dikatakan memiliki reliabilitas untuk dipergunakan karena menghasilkan nilai yang mencerminkan rata-rata harga pasar di wilayah tersebut.

Kertas Kerja Penilaian Wilayah Bandung

Variabel Eksternal	Variabel Properti
Jarak ke CBD terdekat (meter): 428.36	Kondisi Interior: Baik
Lebar Jalan (meter): 10.00	Ukuran_Kamar (m2): 21
Kategori Pesaing (radius 1km): Tinggi >3	Wifi: Tidak Ada
Kategori Kualitas: Baik	Kamar Mandi Dalam: Ada
Jumlah Pengunjung: 2	Amenities Kamar Mandi: Tidak Ada
	Sarapan: Tidak Ada
	Furniture: Ada
	Kulkas: Tidak Ada
	Televisi: Tidak Ada
	Ketel: Tidak Ada
	Jenis Tempat Tidur: Double bed
	Air Conditioner: Ada
Occupancy Rate Kamar per Tahun (%): 50	
<input type="button" value="Hitung Nilai Sewa Kamar"/>	
Nilai sewa kamar per hari Rp 242256.72	

Gambar 9. Hasil Perhitungan Nilai Sewa Kamar Menggunakan Model *Random Forest* Wilayah Bandung (sumber: data sekunder diolah)

KESIMPULAN

Penelitian ini menunjukkan bahwa *machine learning* memiliki potensi besar dalam penilaian properti, menawarkan pendekatan yang lebih andal, skalabel, dan akurat dibandingkan model linear sederhana. Dengan membandingkan tiga algoritma utama (*Ridge Regression*, *Random Forest*, *XGBoost*), penelitian ini menemukan bahwa *Random Forest* memberikan performa terbaik untuk memprediksi harga sewa kamar, terutama di Bandung, dengan R^2 sebesar 0,49 dan NRMSE sebesar 0,13. Hasil estimasi ini memudahkan analisis lebih lanjut dan memungkinkan rekomendasi kebijakan berdasarkan temuan tersebut. Namun, algoritma *machine learning* menawarkan pendekatan alternatif yang lebih andal dalam hal prediksi, meskipun menghadapi tantangan dalam interpretasi hasil. Koefisien yang dihasilkan oleh *machine learning* cenderung sulit diinterpretasikan,

Keterbatasan dalam penelitian ini adalah hasil model dapat bervariasi ketika diterapkan di kota dengan karakteristik data yang berbeda, seperti Bogor dan Jakarta, yang memiliki jumlah data lebih terbatas. Keterbatasan data ini menyebabkan variasi data *training* yang kurang mencakup semua variabel membatasi kemampuan model untuk menangkap kompleksitas penuh dari data. Oleh karena itu, penelitian di masa depan perlu mengumpulkan data yang lebih representatif dari berbagai wilayah, termasuk kategori yang saat ini kurang terwakili, untuk meningkatkan akurasi model.

Penelitian selanjutnya perlu melakukan uji perbandingan antara hasil penilaian sewa secara konvensional dan prediksi *machine learning* menggunakan data yang sama untuk mengevaluasi model secara lebih menyeluruh. Selain itu, penting untuk mengembangkan metode interpretasi yang lebih baik untuk *machine learning*, sehingga hasil prediksi tidak hanya akurat tetapi juga mudah dipahami oleh pengguna dan mendukung pengambilan keputusan dalam konteks properti.

Secara keseluruhan, penelitian ini menghadirkan inovasi dalam teknik penilaian sewa properti Barang Milik Negara (BMN) dengan memanfaatkan *big data* dari data pasar aktual. Sebelumnya, penilaian sewa aset BMN dilakukan secara konvensional dengan membandingkan data objek dengan dua atau lebih objek pembanding menggunakan alat hitung *Ms. Excel*. Pendekatan tersebut memiliki keterbatasan dalam mengolah data dalam jumlah besar dan menghasilkan hasil yang kurang akurat serta memakan waktu. Dalam penelitian ini, metode baru yang diperkenalkan menggunakan analisis dari berbagai algoritma yang tidak hanya analisis regresi saja, dengan memanfaatkan *machine learning* berbasis *artificial intelligence* (AI) sebagai pendekatan analitik data untuk mengidentifikasi pengaruh variabel bebas, seperti karakteristik fisik dan data transaksi, terhadap variabel terikat berupa nilai sewa, sehingga menawarkan gambaran yang lebih relevan dan komprehensif. Lebih jauh, *machine learning* memungkinkan pemrosesan jumlah data yang besar secara cepat dan akurat. Pendekatan ini tidak hanya mampu menghasilkan model untuk klasifikasi tetapi juga prediksi nilai sewa, sehingga meningkatkan efisiensi dan keakuratan proses analisis. Penggunaan *machine learning* dalam pengelolaan aset BMN merupakan kebaruan yang belum pernah diimplementasikan sebelumnya, menjadikannya kontribusi unik dalam tata kelola aset negara.

Penelitian ini juga memberikan rekomendasi untuk penyusunan kebijakan publik yang mendukung adopsi metode penilaian baru melalui *automated valuation model* (AVM). Dengan AVM, proses penilaian diharapkan menjadi lebih cepat, efisien, dan akurat, yang pada akhirnya dapat berdampak positif terhadap optimalisasi Penerimaan Negara Bukan Pajak (PNBP) dari sewa BMN. Integrasi teknologi AI dalam penelitian ini tidak hanya mempercepat proses penilaian tetapi juga memberikan solusi modern dan relevan untuk meningkatkan kualitas pengelolaan aset negara. Penelitian ini diharapkan menjadi dasar untuk pengembangan kebijakan berbasis teknologi, mendukung digitalisasi dan efisiensi dalam pengelolaan aset BMN.

DAFTAR PUSTAKA

- Alpaydin, E. (2020). *Introduction to machine learning* (4th ed.). MIT Press.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. Retrieved from <http://www.spss.com/worldwide>

-
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Dewi, N. K. (2011). Penerapan metode random forest dalam driver analysis. *Forum Statistika dan Komputasi*, 35, 35–43.
- Direktorat Jenderal Kekayaan Negara. (2021). Peraturan Direktur Jenderal Kekayaan Negara Nomor 4 Tahun 2021 tentang Pedoman Teknis Penilaian Sewa Oleh Penilai Pemerintah. Direktorat Jenderal Kekayaan Negara Kementerian Keuangan, Indonesia.
- Direktorat Penilaian Direktorat Jenderal Kekayaan Negara. (2022). Data sewa barang milik negara. Direktorat Jenderal Kekayaan Negara Kementerian Keuangan, Indonesia.
- Gong, H., Sun, Y., Shu, X., & Huang, B. (2018). Use of random forests regression for predicting IRI of asphalt pavements. *Construction and Building Materials*, 189, 890–897. <https://doi.org/10.1016/j.conbuildmat.2018.09.017>
- Ho, W. K. O., Tang, B. S., & Wong, S. W. (2020). Predicting property prices with machine learning algorithms. *Journal of Property Research*. <https://doi.org/10.1080/09599916.2020.1832558>
- Izquierdo-Verdiguier, E., & Zurita-Milla, R. (2020). An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 88, 1–13. <https://doi.org/10.1016/j.jag.2020.102051>
- Kementerian Keuangan. (2020). Peraturan Menteri Keuangan Nomor 115 Tahun 2020 Tentang Pemanfaatan BMN. Kementerian Keuangan, Indonesia.
- Kementerian Keuangan. (2023). Laporan keuangan pemerintah pusat (audited). Kementerian Keuangan, Indonesia.
- Kim, K. (2019). Ridge regression for better usage. Retrieved from <https://towardsdatascience.com/ridge-regression-for-better-usage-2f19b3a202db>
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Nedjati-Gilani, G., Schneider, T., Hall, M., Cawley, N., Hill, I., Ciccarelli, O., Drobnjak, I., Gandini Wheeler-Kingshott, C., & Alexander, D. (2017). Machine learning-based compartment models with permeability for white matter microstructure imaging. *NeuroImage*, 150, 119–135. <https://doi.org/10.1016/j.neuroimage.2017.02.013>
- Phan, T. D. (2018). Housing price prediction using machine learning algorithms: The case of Melbourne City, Australia. *2018 International Conference on Machine Learning and Data Engineering (iCMLDE)*, 35–42. <https://doi.org/10.1109/iCMLDE.2018.00017>
- Roser, M., Ritchie, H., & Ortiz-Ospina, E. (2019). World population growth. Retrieved January 20, 2020, from *Our World in Data*: <https://ourworldindata.org/world-population-growth>
- Sangha, A. (2021). Property valuation by machine learning for the Norwegian real estate market. *SAGE Open Research*. <https://doi.org/10.14293/S2199-1006.1.SOR-PP0TP9I.v1>
- Suhardi, G. (2009). *Analisis kepadatan penduduk di Kecamatan Delanggu Kabupaten Klaten 2006* (Undergraduate thesis, Universitas Sebelas Maret, Surakarta).
- Susanto, H., & Sudiyanto. (2014). Penerapan data mining untuk prediksi penjualan wallpaper menggunakan algoritma C4.5. *Jurnal Pendidikan Vokasi*, 4(2), 225–232.