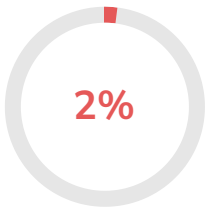
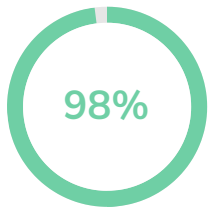
Plagiarism Scan Report

Report Generated on: Aug 29,2022



Plagiarised



Unique

Total Words:	**995**
Total Characters:	**7733**
Plagiarized Sentences:	**0.94**
Unique Sentences:	**46.06 (98%)**

Content Checked for Plagiarism

Penerapan Random Forest Regression Untuk Memprediksi Harga Jual Rumah Dan Cosine Similarity Untuk Rekomendasi Rumah Pada Provinsi Jawa Barat

Ega Sri Lestari1, Ida Astuti2

Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadarma1,2
Jl. Margonda Raya 100, Depok, Jawa Barat, 16424
egasrilestari@student.gunadarma.ac.id1, astuti@staff.gunadarma.ac.id2

(received= dd-mm-yy, revised= dd-mm-yy, accepted= dd-mm-yy (diisi oleh editor))

ABSTRACT
This study aims to implement a machine learning algorithm, namely Random Forest Regression in predicting house prices and the Cosine Similarity algorithm in providing house recommendations. The data is taken using web scrapping technique, then the data will be processed using the CRISP-DM (Cross-Industry Standard Process dor Data Mining) method with business understanding, data understanding, data preparation, modeling, evaluation and deployment stages. The results of the accuracy of the prediction process with tuning parameters are 85.29%, while the results of the recommendation accuracy are 89.99% on the test data. The research has succeeded in implementing a model in the form of a website that can be used by users in searching for house price needs in the province of West Java, then recommendations are given by switching the link to the rumah123.com website for more complete information. The website that has been built has been tested by inferential testing to get a precision value obtained by the system 75%, recall 100%, system accuracy 80%, while the f-measure is 86%, while the recommendation system gets a precision value obtained by the system 78%, recall 100%, 80% system accuracy, while the f-measure is 88%. User acceptance testing on the website gets a percentage of 89.29%.

Keyword: Cosine Similarity, Prediction of house prices, Random Forest Regression., Recommendation of house prices.

ABSTRAK
Penelitian ini bertujuan untuk mengimplementasikan sebuah algoritma machine learning yaitu Random Forest Regression dalam memprediksi harga rumah dan algoritma Cosine Similarity dalam memberikan rekomendasi rumah. Data yang diambil menggunakan teknik web scrapping, lalu data tersebut akan di olah menggunakan metode CRISP-DM (Cross-Industry Standard Process dor Data Mining) dengan tahap business understanding, data understanding, data preparation, modeling, evaluation dan deployment. Hasil akurasi dari proses prediksi dengan tuning parameter sebesar 85,29%, sedangkan hasil akurasi rekomendasi mendapatkan hasil 89,99% pada data uji. Penelitian berhasil mengimplementasikan model berupa website yang dapat digunakan oleh pengguna dalam mencari kebutuhan harga rumah di daerah provinsi Jawa Barat kemudian rekomendasi diberikan dengan peralihan link menuju website rumah123.com untuk informasi lebih lengkap. Website yang sudah dibangun telah diuji dengan pengujian inferensial mendapatkan nilai precision yang didapatkan oleh

sistem 75%, recall 100%, akurasi sistem 80%, sedangkan f-measure 86%, sedangkan pada sistem rekomendasi mendapatkan nilai precision yang didapatkan oleh sistem 78%, recall 100%, akurasi sistem 80%, sedangkan f-measure 88%. Pengujian user acceptance pada website mendapatkan persentase sebesar 89.29%.

Kata Kunci: Cosine Similarity, Prediksi Harga Rumah, Random Forest Regression, Rekomendasi Harga Rumah.

I. PENDAHULUAN

Provinsi Jawa Barat merupakan sebuah provinsi yang memiliki jumlah penduduk terbanyak dibandingkan dengan provinsi lain di Indonesia. Terbukti bahwa provinsi Jawa Barat memiliki tingkat kepadatan penduduk yang cukup tinggi berdasarkan data pertumbuhan penduduk di Pulau Jawa dimana provinsi Jawa Barat berada pada peringkat pertama yang memiliki jumlah pertumbuhan penduduk paling banyak pada tahun 2016 hingga tahun 2020 [1]. Pertumbuhan penduduk ini mengakibatkan masyarakat harus memiliki tempat tinggal yang layak dengan harga yang terjangkau sesuai dengan kebutuhannya masing-masing.

Rumah merupakan salah satu dari banyaknya kebutuhan primer bagi setiap orang khususnya keluarga dalam mendapatkan sebuah tempat tinggal pribadi demi melindungi penghuni rumah dari berbagai kondisi alam yang ada di sekitarnya seperti hujan, panas terik matahari, angin kencang, dan sebagainya [2]. Aplikasi prediksi dengan memberikan rekomendasi sangat cocok digunakan pada masyarakat untuk mendapatkan rumah yang diinginkan dengan cepat dan mudah.

Secara sederhana Machine Learning ini adalah salah satu cabang ilmu kecerdasan buatan, khususnya dapat mempelajari tentang bagaimana komputer mampu belajar dari data untuk meningkatkan kecerdasannya [3]. Pada tahun 2020, telah dilakukan sebuah penelitian penggunaan algoritma Random Forest Regression untuk memprediksi harga sewa apartment yang memiliki tingkat akurasi sebesar 92% dengan peningkatan menggunakan tuning parameter sebesar 0.89% [4]. Pada tahun 2021, telah dilakukan penelitian dengan menggunakan algoritma Linear Regression untuk memprediksi harga rumah di daerah Tebet. Hasil akurasi yang didapatkan sebesar 72% tanpa melakukan tuning parameter untuk meningkatkan akurasi [2].

Memberikan sistem rekomendasi kepada pengguna agar memudahkan dalam melakukan pencarian referensi rumah. Pada tahun 2020, Penelitian dalam menggunakan Cosine Similarity dalam mencari keyword pada sesuatu yang hilang menggunakan metode TF-IDF menghasilkan akurasi sebesar 88% [5].

Berdasarkan beberapa penelitian yang telah dilakukan sebelumnya, peneliti tertarik untuk melakukan penelitian prediksi harga rumah di provinsi Jawa Barat menggunakan algoritma Random Forest Regression serta memberikan rekomendasi referensi rumah menggunakan algoritma Cosine Similarity. Alasan menggunakan algoritma Random Forest Regression ini dikarenakan pada beberapa penelitian terkait algoritma ini memberikan hasil yang paling bagus untuk mengatasi permasalahan prediksi atau regresi lain.

II. Metodologi Penelitian

Metode penelitian ini akan menggunakan salah satu metodologi data science yaitu CRISP-DM (CrossIndustry Standard Process for Data Mining).

Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1 di atas penjabaran tahap penelitian yang akan dilakukan sebagai berikut:

Business Understanding

Masalah yang akan diselesaikan pada penelitian ini adalah mewujudkan hasil prediksi rumah berdasarkan input pengguna serta memberikan rekomendasi rumah yang berkaitan dengan profil pengguna tersebut. Metode yang cocok digunakan untuk menyelesaikan masalah ini menggunakan algoritma regresi yang dapat memprediksi nilai terhadap inputan yang diberikan serta menggunakan algoritma perhitungan jarak agar menghasilkan rekomendasi yang dapat disesuaikan dengan kemiripan data input.

Data Understanding

Pada tahap kedua akan melakukan proses untuk pengambilan data penelitian yaitu menggunakan teknik web scrapping pada website https://www.rumah123.com/jual/jawa-barat/rumah/. Tahap validasi ini diantaranya adalah mengganti nama kolom pada data, mengganti tipe data, membuang data yang tidak diperlukan. Proses selanjutnya adalah menelaah data yang bertujuan untuk menganalisa data secara

statistika deskriptif agar dapat melihat karakteristik data, kemudian melakukan pengecekan terhadap korelasi antar kolom untuk melihat keterhubungan kolom yang saling mempengaruhi. Tahap terakhir dalam data understanding adalah eksplorasi data dapat dilakukan dengan berbagai macam seperti membentuk pivot tabel, visualisasi diagram dan sebagainya.

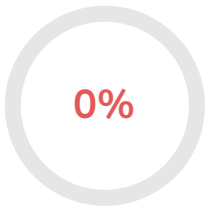
Data Preparation
Melakukan tahap seleksi data untuk menentukan kolom yang memiliki pengaruh paling besar dengan target penelitian. Pembersihan data bertujuan untuk meminimalkan noise/data yang tidak lengkap sehingga data menjadi bersih dari bias/kesalahpahaman data

Petunjuk Penulisan Jurnal Ilmiah FIFO (Font size14, center, Bold)
(received= dd-mm-yy, revised= dd-mm-yy, accepted= dd-mm-yy (diisi oleh editor))
http://mukhyi.staff.gunadarma.ac.id/Downloads/files/104121/Draft+Jurnal+Ilmiah+FIFO+AditiaDewiMukhyi.pdf

100%

Plagiarism Scan Report

Report Generated on: Aug 29,2022



Plagiarised

Unique

Total Words:	**985**
Total Characters:	**7146**
Plagiarized Sentences:	**0**
Unique Sentences:	**43 (100%)**

Content Checked for Plagiarism

Pembersihan data ini dimulai dari pengecekan nilai kosong/missing value, lalu mengecek data duplikat dan yang terakhir adalah pengecekan outlier. Salah satu cara yang dilakukan untuk mengecek outlier pada data dengan metode Interquartile Range (IQR) dengan rumus sebagai berikut:
$IQR = Kuartil3 - Kuartil1$ (1)
Lalu cek nilai outlier dengan syarat data < Kuartil 1 – 1.5 * IQR atau data > Kuartil 3 + 1.5 * IQR.
Konstruksi data bertujuan untuk menambahkan fitur, normalisasi data dan melakukan transformasi data non-numerik menjadi numerik. Pada proses normalisasi akan menggunakan metode MinMaxScaler dengan rumus sebagai berikut:
(2)
Keterangan:
Newdata = Data baru hasil normalisasi.
Data = Data yang akan dinormalisasi.
Min = Nilai terkecil dari data yang akan dinormalisasi.
Max = Nilai terbesar dari data yang akan dinormalisasi.
Newmin = Batas nilai terkecil normalisasi.
Newmax = Batas nilai terbesar normalisasi.

Modeling
Pemodelan data latih dengan algoritma Random Forest Regression untuk memprediksi harga rumah dan Cosine Similarity untuk rekomendasi rumah berdasarkan hasil prediksi. Pemisahan data yang akan dilakukan menggunakan 3 skenario diantaranya skenario pertama 80% data latih dan 20% uji, skenario kedua 75% data latih dan 25% data uji, skenario ketiga 70% data latih dan 30% data uji.
Random Forest juga merupakan algoritma berbasis “Pohon” yang menggunakan fitur kualitas dari beberapa pohon keputusan (Decision Tree) untuk membuat keputusan agar model cukup untuk mempelajari fitur-fitur dari data [6]. Pembentukan akar pohon akan diambil berdasarkan nilai terkecil splitting criteria pada masing-masing fitur yang digunakan. Pada masalah regresi menggunakan Random Forest disarankan untuk menggunakan perhitungan MSE dalam menentukan splitting criteria pohon dengan menggunakan rumus sebagai berikut:
(3)
Keterangan:
$MSEn$= Nilai MSE pada pohon ke-n.
N = Jumlah sampel pada pohon ke-n.
Yi = Nilai sampel ke-I pada pohon ke-n.
$\bar{Y}n$ = Nilai rata-rata sampel pohon ke-n.
Prediksi pada kasus regresi diambil dari nilai rata-rata dari setiap pohon Rumus untuk menghitung nilai ratarata seluruh prediksi pohon sebagai berikut:
ˆ1 N ˆ(4)
Keterangan:
Hasil prediksi akhir.
$Ntree$ = Total jumlah pohon pada Random Forest.
asil prediksi pohon ke-n.
Pengujian performance dari model yang telah dibentuk akan dibandingkan dengan menggunakan tuning parameter dimana terdapat 2 model yang akan dibandingkan yaitu model dengan parameter default dan

model dengan parameter pilihan dari tuning. Metode tuning parameter yang akan digunakan adalah Grid Search CV untuk data pada setiap fold akan dikombinasikan dengan parameter yang digunakan dan grid search akan memilih kombinasi terbaik berdasarkan nilai CV Score yang tertinggi [7].
Cosine Similarity ini akan mengukur sudut antara perkalian dua vektor yang diproyeksikan dalam ruangan multidimensi. Nilai Cosine Similarity yang paling tinggi akan menjadi hasil rekomendasi kepada user [8]. Rumus yang akan digunakan untuk perhitungan Cosine Similarity sebagai berikut:
(5)
Keterangan:
Q = Vektor Q, dibandingkan kemiripan datanya.
D = Vektor D, dibandingkan kemiripan datanya.
$Q \cdot D$ = Dot product antara vektor Q dan D.
$|Q|$ = Panjang vektor Q.
$|D|$ = Panjang vektor D.
= Cross product antara vektor Q dan vektor D.
= Term ke-y yang terdapat pada data WQ.
= Term ke-y yang terdapat pada dokumen Wi.

Evaluation
Tahap evaluasi akan melakukan pengukuran akurasi data terhadap hasil prediksi dan rekomendasi yang diberikan menggunakan data testing pada masing-masing model. Sebuah model Machine Learning dikatakan baik harus memberikan hasil output yang akurat dalam rangka untuk pengambilan keputusan berdasarkan hasil yang diberikan [9]. Pengukuran yang akan dilakukan menggunakan metrik MSE (Mean Squared Error), MAE (Mean Absolute Error) dan R-Squared.
MSE dapat dikatakan sebagai rata-rata kuadrat kesalahan yang dapat dihitung dengan menjumlahkan semua kesalahan atau error hasil prediksi atau peramalan pada setiap data input yang dimasukan kemudian dikuadratkan dan membaginya dengan jumlah periode peramalan [10]. Rumus untuk menghitung MSE sebagai berikut:
(6)
Keterangan:
Xi = Data aktual pada periode ke-i.
Fi = Nilai hasil ramalan atau prediksi pada periode ke-i. n = Banyaknya data sampel.
Pada pengukuran model MAE setiap selisih error akan diambil nilai mutlak nya untuk selanjutnya di jumlahkan. Rumus untuk menghitung MAE sebagai berikut:
(7)
Keterangan:
n = Banyaknya data sampel.
= Nilai aktual untuk sample j.
Nilai prediksi untuk sample j.
Pengukuran model dengan R-squared adalah proporsi variasi dalam variabel dependen yang dapat dijelaskan oleh variabel-variabel independen nya - sebagai berikut:
(8)
Keterangan:
= Data uji.
= Data uji hasil prediksi.
= Rata-rata data uji.

Deployment
Tahap deployment dimana akan mengimplementasikan hasil model yang telah disimpan dan hasil rekomendasi ke dalam website menggunakan framework yang disediakan Python yaitu flask. Tahap terakhir pada penelitian adalah pengujian website prediksi harga rumah yang akan menggunakan pengujian inferensial dan pengujian user acceptance.

Gambar 2. Use Case Website
Berdasarkan Gambar 2 di atas merupakan use case dari website yang akan dibangun, user yang dapat menggunakan website ini hanya pengguna. Kegiatan yang dapat dilakukan oleh pengguna website meliputi melihat beranda yang menampilkan informasi website dan tipe-tipe rumah, kemudian dapat melihat halaman tentang yang menampilkan deskripsi singkat pembuat website. Pengguna juga dapat memasukkan data prediksi yang akan menghasilkan nilai prediksi dan rekomendasi dari implementasi model. Kegiatan lain yang dapat dilakukan pengguna adalah melihat hasil prediksi dan hasil

rekomendasi.

Gambar 3. Activity Diagram Website

Berdasarkan Gambar 3 di atas aktifitas pertama yang akan dilakukan oleh pengguna adalah membuka halaman beranda website prediksi, lalu website akan menampilkan halaman beranda. Apabila pengguna tidak membuka halaman beranda dan membuka halaman prediksi maka website akan menampilkan halaman prediksi, lalu pengguna diharuskan untuk memasukkan kriteria rumah sehingga website akan menampilkan hasil prediksi. Jika data yang dimasukkan pengguna mendapatkan hasil rekomendasi maka website akan menampilkan data rekomendasi, namun apabila data yang dimasukkan tidak mendapatkan rekomendasi maka tidak akan menampilkan hasil rekomendasi. Apabila pengguna sudah melihat hasil prediksi dan rekomendasi maka pengguna juga dapat melihat informasi lengkap mengenai rumah rekomendasi yang diberikan jika rumah masih tersedia maka website rumah123.com akan menampilkan informasi lebih lengkap terkait rumah tersebut seperti detail dari lokasi rumah, gambar dari rumah yang ditawarkan dan informasi lainnya mengenai rumah yang ditawarkan, apabila rumah sudah tidak tersedia maka website rumah123.com akan menampilkan halaman error 404 dikarenakan ada kemungkinan bahwa rumah sudah tidak di iklankan pada website rumah123.com.

Plagiarism Scan Report

Report Generated on: Aug 29,2022

0%	Plagiarised	
100%	Unique	
Total Words:	990	
Total Characters:	7134	
Plagiarized Sentences:	0	
Unique Sentences:	45 (100%)	

Content Checked for Plagiarism

III. Hasil dan Pembahasan
Hasil pengumpulan data dari proses web scrapping pada website rumah123.com mendapatkan data sebanyak 1004 baris dan 9 kolom yang diambil dari tanggal 1 Maret 2022 hingga 5 Maret 2022 hanya wilayah provinsi Jawa Barat.

Tabel 1. Hasil Pengumpulan Data
no. nama daerah harga kamar tidur kamar mandi garasi luas tanah luas bangunan link
1 Cluster Sawangan. Depok 485 Juta 2 1 1 60 45 /properti/depok/hos9945070/
2 Rumah Mewah Depok Depok 1,42
Miliar 4 3 2 93 130 /properti/depok/hos9612079/
: : : : : : : : : : : : : : : : : :
1003 New Diamond Depok 800 Juta 2 2 1 97 45 /properti/depok/hos9748561/
1004 Hunian Cantik Depok Depok 950 Juta 3 3 2 95 100 /properti/depok/hos9943216/

Berdasarkan Tabel 1 dapat dilihat bahwa pada kolom harga masih menggunakan tipe data string sehingga harus diubah ke dalam bentuk integer menggunakan metode Regex dengan syarat kata “Juta” akan dikalikan dengan 1E6 sedangkan kata “Miliar” akan dikalikan dengan 1E9. Pada proses validasi ini akan melakukan penghapusan terhadap data yang tidak dibutuhkan seperti mengecek data dengan nama rumah yang mempunyai kata “kos”, “kost”, “kostan” data tersebut akan dihapus karena akan memberikan efek ketidakstabilan pada data. Lakukan penghapusan terhadap data yang tidak diperlukan kemudian dilakukan cek terhadap jumlah data yang sudah bersih dari penyimpangan sehingga jumlah data sekarang sebanyak 994 baris data dengan 9 kolom.

Gambar 4. Hasil Korelasi Data
Berdasarkan Gambar 4 di atas terlihat bahwa harga rumah memiliki korelasi paling tinggi dengan luas bangunan yaitu 0.87 yang berarti jika harga rumah mengalami peningkatan maka hal tersebut sangat dipengaruhi oleh luas bangunan. Pada korelasi antara harga rumah dengan jumlah garasi adalah korelasi yang kecil akan tetapi memiliki nilai yang positif berarti harga rumah sulit untuk di pengaruhi oleh jumlah garasi dari rumah tersebut.
Proses seleksi data menghasilkan hasil yang sama dengan korelasi sehingga menyatakan bahwa fitur kolom luas bangunan memiliki kontribusi yang paling berpengaruh dalam memprediksi harga rumah.

Gambar 5. Cek Nilai Hilang
Berdasarkan Gambar 5 di atas pada proses pembersihan data akan melakukan perbaikan, penghapusan dan pengabaian dari noise yang ada pada data. Semua jumlah nilai pada setiap atribut memiliki jumlah yang sama yaitu masing-masing berjumlah 994 data sehingga tidak dibutuhkan untuk melakukan penanganan terhadap nilai hilang. Jika pada data terdapat nilai yang hilang maka harus segera ditangani sesuai dengan karakteristik data tersebut. Pada setiap penanganan nilai hilang memiliki kriteria yang dapat disesuaikan dengan data.

Gambar 6. Cek Data Duplikat
Berdasarkan Gambar 6 di atas jika sudah selesai mengecek nilai kosong pada data maka lakukan pengecekan terhadap data duplikat, dari hasil yang telah dilakukan terdapat 4 data duplikat sehingga data duplikat ini dapat dihapus sehingga jumlah data tanpa duplikat sebanyak 990 baris dengan 9

kolom. Metode yang akan digunakan untuk mengecek nilai outlier adalah IQR dan boxplot.

Gambar 7. Outlier Pada Seluruh Data
Berdasarkan Gambar 7 di atas kolom yang mempunyai outlier dengan jumlah terbanyak dari seluruh data yang digunakan. Jika di hitung menggunakan rumus IQR pada kolom luas tanah memiliki outlier paling banyak yaitu sebanyak 85 data, data yang memiliki outlier akan dihapus barisnya sehingga data yang sudah bersih siap untuk di modeling memiliki 848 baris dan 9 kolom.
Proses selanjutnya adalah melakukan konstruksi data dengan transformasi kolom daerah menjadi tipe data integer. Transformasi ini akan menggunakan function LabelEncoder pada Python. Proses transformasi ini sangat penting dilakukan karena Machine Learning tidak dapat mengolah data dengan sifat string atau object.

Gambar 8. Hasil Pelabelan Data
Berdasarkan Gambar 8 di atas dapat dilihat bahwa sekarang data kota sudah berubah menjadi angka yang diurutkan berdasarkan abjad daerah tersebut. Proses normalisasi menggunakan MinMaxScaler untuk mengubah rentang pada data sehingga semua kolom pada saat modeling akan memiliki rentang yang sama yaitu 0-1.

Gambar 9. Hasil Normalisasi Data
Berdasarkan Gambar 9 di atas hasil data yang telah dinormalisasi sudah memiliki rentang yang sama antar kolom. Jika sudah melakukan proses data konstruksi maka data sudah siap untuk dimasukan ke dalam tahap model dengan algoritma Random Forest Regression.
Tabel 2. Hasil Pemisahan Data
Kriteria Data Latih Data Uji
80:20 659 165
75:25 618 206
70:30 577 247

Berdasarkan Tabel 2 di atas masing-masing skenario akan dilakukan pelatihan dengan algoritma Random Forest Regression dengan parameter yang sama. Pada tahap modeling algoritma Random Forest Regression hanya menggunakan parameter n_estimators, sedangkan untuk parameter lain di gunakan secara default.
forest80 = RandomForestRegressor(n_estimators=300, random_state=42) forest80.fit(X_train80, y_train80) forest75 = RandomForestRegressor(n_estimators=300, random_state=42) forest75.fit(X_train75, y_train75) forest70 = RandomForestRegressor(n_estimators=300, random_state=42) forest70.fit(X_train70, y_train70)
Pada code di atas merupakan proses melakukan pelatihan terhadap data training dengan algoritma Random Forest Regression menggunakan Python. Parameter yang ditentukan hanya jumlah pohon yang terbentuk yaitu 300 sedangkan parameter lain seperti bootstrap, max_features dan lainnya ditentukan secara default. Pada setiap skenario akan dilakukan pelatihan dengan algoritma Random Forest Regression dengan parameter yang sama. Pada tuning parameter ini akan memanfaatkan library GridSearchCV dalam memberikan parameter terbaik.

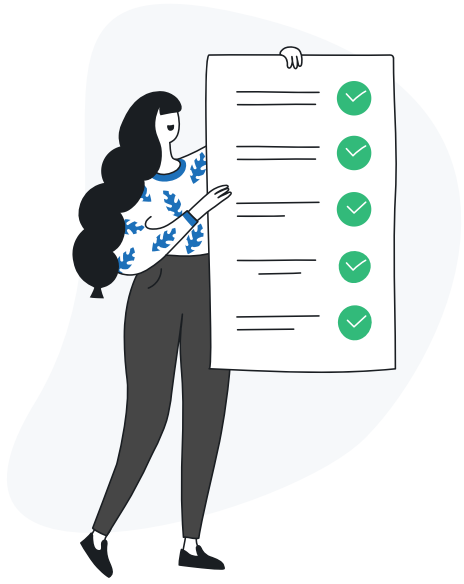
Tabel 3. Akurasi Model 3 Skenario
Skenario Sebelum Tuning Setelah Tuning
80:20 78.93 80.92
75:25 80.30 85.29
70:30 79.17 83.20

Berdasarkan Tabel 3 di atas akurasi data uji tertinggi didapatkan oleh kriteria 75% data latih dan 25% data uji dengan hasil prediksi menggunakan parameter metode Grid Search CV naik sekitar 4.99%.

Gambar 10. Hasil Algoritma Random Forest Regression Setelah Tuning Parameter
Berdasarkan Gambar 10 di atas merupakan hasil pohon yang berhasil dibentuk oleh model dengan split data 75% data latih dan 25% data uji setelah menggunakan parameter gridsearchCV. Pada proses pemodelan algoritma Cosine Similarity akan menggunakan data split dengan kriteria 75% data latih dan 25% data uji. Pembentukan data untuk memberikan hasil rekomendasi berdasarkan nilai Cosine Similarity antara data latih dengan data input yang dimasukan pengguna. Proses content based filtering sebagai sebuah metode tambahan untuk pelengkap dari hasil rekomendasi rumah yang diberikan.

Metode filtering tambahan ini akan memberikan efek untuk memperkecil lingkup rekomendasi sehingga hasil rekomendasi yang diberikan akan semakin akurat mendekati profil pengguna. Fitur yang akan di gunakan adalah 3 fitur yang memiliki kontribusi penting bagi pengguna dalam memilih harga rumah yaitu daerah, harga dan luas tanah.

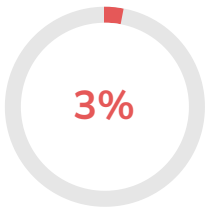
Gambar 11. Proses Cosine Similarity



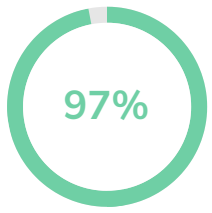
No Plagiarism Found

Plagiarism Scan Report

Report Generated on: Aug 29,2022



Plagiarised



Unique

Total Words:	**994**
Total Characters:	**7056**
Plagiarized Sentences:	**1.62**
Unique Sentences:	**52.38 (97%)**

Content Checked for Plagiarism

Berdasarkan Gambar 11 di atas merupakan proses perolehan data prediksi dan rekomendasi yang akan diberikan. Dari data input pengguna maka akan diberikan hasil harga prediksi rumah berdasarkan kriteria tersebut. Harga prediksi ini akan di lakukan proses dengan algoritma Cosine Similarity dan menggunakan filtering 3 faktor tambahan yaitu harga, kota dan luas tanah. hasil yang diberikan oleh perhitungan Cosine Similarity ini cukup dekat dengan data input yang dimasukan oleh pengguna. Terlihat dari kota yang dipilih, luas bangunan yang hampir mendekati data input, dan sebagainya. Proses terakhir mengecek akurasi dari model yang sudah dibentuk dengan pemisahan 75% data latih dan 25% data uji pada model prediksi dan model rekomendasi. Metrik yang akan digunakan adalah MSE, MAE, dan R-Squared menggunakan bahasa pemrograman Python.

Gambar 12. Hasil Evaluasi Model
Berdasarkan Gambar 12 di atas dengan menggunakan Python pada seluruh dataset yang digunakan untuk pelatihan random forest setelah tuning parameter mengalami penurunan error baik pada metrik MSE maupun MAE. Pada sistem rekomendasi dengan menggunakan algoritma Cosine Similarity mendapatkan hasil 6.26 pada MSE dan 2.05 pada MAE sehingga dapat dikatakan bahwa rekomendasi yang dihasilkan cukup baik. Hasil evaluasi pada algoritma Random Forest Regression mengalami peningkatan yang cukup baik setelah melakukan proses tuning parameter dengan hasil sebesar 85.29% dan pada algoritma Cosine Similarity sebesar 89.99%. Proses terakhir pada penelitian ini adalah deployment yang akan melakukan implementasi model yang sudah disimpan menjadi sebuah sistem berbasis website agar dapat diakses dan digunakan oleh masyarakat dalam membantu menemukan rumah yang diinginkan. Pada proses pembuatan website sebagai implementasi dari hasil analisis ini akan menggunakan framework Flask yang ada pada bahasa pemrograman Python sebagai backend, lalu menggunakan HTML dan CSS sebagai frontend dari website yang akan dibangun.

Gambar 13. Tampilan Halaman Beranda
Berdasarkan Gambar 13 di atas merupakan tampilan halaman beranda pada website prediksi. Pada halaman ini akan menampilkan informasi tentang tipe-tipe rumah mulai dari rumah 1 lantai hingga 5 lantai.

Gambar 14. Tampilan Halaman Tentang
Berdasarkan Gambar 14 di atas pada tampilan halaman tentang akan menampilkan informasi singkat mengenai deskripsi penulis dan deskripsi penggunaan website prediksi untuk pada pengguna aplikasi.

Gambar 15. Tampilan Halaman Prediksi
Berdasarkan Gambar 15 di atas merupakan hasil tampilan halaman prediksi yang bertujuan untuk memberikan input kriteria rumah yang akan diprediksi.

Gambar 16. Tampilan Halaman Hasil
Berdasarkan Gambar 16 di atas merupakan tampilan hasil prediksi harga rumah dan halaman ini akan ditampilkan informasi kriteria yang dipilih beserta hasil prediksi harga rumah. Setelah menyelesaikan hosting maka website sudah dapat diakses secara online oleh masyarakat dengan alamat url: https://websiteprediksihargarumahjabar.herokuapp.com/.
Pengujian inferensial akan melihat kinerja model yang diberikan setelah dilakukan deployment. Terdapat

10 data uji yang akan digunakan untuk menguji kinerja prediksi model dan rekomendasi model yang diberikan. Tabel 4. Data Pengujian Inferensial Model Prediksi
data uji ke- harga prediksi selisih label prediksi label aktual
data uji ke-1 365000000 355726000 9274000 relevan relevan
data uji ke-2 550000000 484598200 65401800 tidak relevan tidak relevan
data uji ke-3 567000000 591194300 24194300 relevan relevan
data uji ke-4 380000000 352048500 27951500 relevan relevan
data uji ke-5 314000000 320622500 6622500 relevan tidak relevan
data uji ke-6 1200000000 923332500 276667500 tidak relevan tidak relevan
data uji ke-7 489000000 450581500 38418500 relevan relevan
data uji ke-8 412000000 448487500 36487500 relevan relevan
data uji ke-9 457000000 482225500 25225500 relevan relevan
data uji ke10 350000000 303722500 46277500 relevan tidak relevan
Berdasarkan Tabel 4 di atas pada kolom selisih di dapatkan dari hasil nilai absolute pengurangan dari harga asli dengan harga prediksi yang diberikan. Dari kolom selisih tersebut maka akan dihitung nilai rata-rata selisih harga yang diberikan antara harga asli dengan harga prediksi.
Pada label prediksi diberi sebuah kondisi:
• Relevan: apabila nilai selisih <= rata-rata selisih.
• Tidak relevan: apabila nilai selisih > rata-rata selisih.
Pada label aktual diberi sebuah kondisi berdasarkan beberapa asumsi pengguna yang menggunakan sistem:
• Relevan: apabila nilai selisih < 40.000.000.
• Tidak relevan: apabila nilai selisih > 40.000.000.

Gambar 17. Confusion Matrix Sistem Prediksi

0.86
Tabel 5. Data Pengujian Inferensial Model Rekomendasi data uji ke- total rekomendasi label rekomendasi label aktual data uji ke-1 5 relevan relevan data uji ke-2 5 relevan tidak relevan data uji ke-3 5 relevan tidak relevan
data uji ke-4 5 relevan relevan
data uji ke-5 5 relevan relevan
data uji ke-6 1 relevan relevan
data uji ke-7 4 relevan relevan
data uji ke-8 5 relevan relevan
data uji ke-9 4 relevan relevan
data uji ke-10 0 tidak relevan tidak relevan

Berdasarkan Tabel 5 merupakan data label yang diberikan sistem kepada pengguna. Total rekomendasi yang diberikan oleh sistem maksimal adalah 5 data rekomendasi.
Pada label rekomendasi diberikan sebuah kondisi:
• Label relevan: apabila rekomendasi yang diberikan sistem terhadap profil pengguna jika terdapat minimal 1 data rekomendasi.
• Label tidak relevan: apabila tidak ada data rekomendasi yang diberikan oleh sistem.
Pada label aktual diberikan sebuah kondisi sebagai berikut:
• Label relevan: jika data rekomendasi yang diberikan setidaknya memiliki 4 kriteria yang sesuai dengan profil pengguna.
• Label tidak relevan: apabila hasil rekomendasi yang diberikan tidak sesuai dengan profil pengguna.

Gambar 18. Confusion Matrix Sistem Rekomendasi

0.88
Pengujian User acceptance dilakukan untuk mengetahui tanggapan dari pengguna pada website yang sudah dibuat. Pengisian kuesioner ini dilakukan dengan 20 responden untuk mengukur kepuasan pengguna terhadap hasil prediksi dan rekomendasi website.

Tabel 6. Hasil Pengujian User acceptance

Penilaian Skor

Pertanyaan Jumlah
SS S N TS STS SS S N TS STS

Apakah menu-menu pada website mudah di
pahami? 12 6 2 0 0 60 24 6 0 0 90
Apakah performa website baik dalam memberikan hasil harga prediksi? 11 5 4 0 0 55 20 12 0 0 87
Apakah website dapat memberikan hasil rekomendasi yang akurat?
Apakah fitur-fitur inputan seperti pemilihan 10 7 3 0 0 50 28 9 0 0 87
daerah, luas tanah dan lain-lain pada website sudah baik? 15 5 0 0 0 75 20 0 0 0 95
Apakah website sudah sesuai dengan
kebutuhan untuk melakukan prediksi dan 11 7 2 0 0 55 28 6 0 0 89 memberikan rekomendasi harga
rumah?
Apakah hasil rekomendasi yang ditampilkan
oleh website sudah sesuai dengan keinginan

LAMPIRAN
SS S N TS STS SS S N TS STS
http://repository.unmuhjember.ac.id/6391/10/lampiran.pdf
100%

pengaruh manajemen kegiatan ekstrakurikuler pramuka ...
15 5 0 0 0 75 20 0 0 0 95
https://www.academia.edu/32857318/PENGARUH_MANAJEMEN_KEGIATAN_EKSTRAKURIKULER_PRAMUKA_TERHADAP_PEMBENTUKAN_AKHLAKUL_KARIMAH_SISWA_DI_SMPN_3_TENGGARONG
100%

Plagiarism Scan Report

Report Generated on: Aug 29,2022

0%	100%	Total Words: 613
Plagiarised	Unique	Total Characters: 4693
		Plagiarized Sentences: 0
		Unique Sentences: 25 (100%)

Content Checked for Plagiarism

Apakah link informasi selengkapnya pada
10 7 3 0 0 50 28 9 0 0 87 website berjalan dengan baik?

Berdasarkan Tabel 6 di atas kolom skor didapatkan dari hasil perkalian penilaian dengan bobot yang di tentukan sesuai dengan label. Kolom jumlah didapatkan dari hasil penjumlahan skor observasi hasil kuesioner yang sudah di beri bobot masing-masing label.

Skor yang diharapkan pada pengujian ini adalah 7 × 100 = 700, nilai 100 didapatkan karena pada pengujian bobot tertinggi adalah label SS dengan nilai 5 kemudian di kalikan dengan 20 responden jadi 100 adalah nilai tertinggi.

dilihat bahwa 20 pengguna website prediksi menilai bahwa hasil yang diberikan oleh aplikasi cukup mudah digunakan dan bermanfaat untuk membantu pengguna dalam menemukan rekomendasi dan prediksi harga rumah yang diinginkan.
IV. Kesimpulan
Pada proses rekomendasi menggunakan algoritma Cosine Similarity dengan metode content based filtering pada fitur daerah, luas tanah, dan harga rumah. Metode yang digunakan adalah CRISP-DM (Cross-Industry Standard Process dor Data Mining) dengan tahap business understanding, data understanding, data preparation, modeling, evaluation dan deployment. Proses prediksi dilakukan menggunakan algoritma Random Forest Regression dengan akurasi tertinggi terdapat pada pemilihan 75% data latih dan 25% data uji. Berdasarkan hasil evaluasi model yang telah dilakukan pelatihan mendapatkan hasil akurasi prediksi dengan 848 data bersih menggunakan tuning parameter sebesar 85,29%, sedangkan pada evaluasi rekomendasi mendapatkan hasil 89,99%.
Pada hasil uji website menggunakan inferensial sistem prediksi mendapatkan nilai precision yang didapatkan oleh sistem 75%, recall 100%, akurasi sistem 80%, sedangkan f-measure 86%. Sehingga dapat dikatakan bahwa model telah bekerja dengan baik dalam memberikan hasil prediksi harga dengan menggunakan data uji yang diberikan. Pada sistem rekomendasi uji inferensial mendapatkan nilai precision yang didapatkan oleh sistem 78%, recall 100%, akurasi sistem 80%, sedangkan f-measure 88%. Sehingga dapat dikatakan bahwa model telah bekerja dengan baik dalam memberikan hasil rekomendasi rumah dengan menggunakan data uji yang diberikan. Pada pengujian user acceptance website mendapatkan persentase sebesar 89.29% disimpulkan bahwa pengguna mudah menggunakan website dan dapat bermanfaat untuk memberikan rekomendasi dan prediksi harga rumah di provinsi Jawa Barat.
Daftar Pustaka
[1] BPS. (2022, Maret). Jumlah Penduduk Hasil Proyeksi Menurut Provinsi dan Jenis Kelamin (Ribu Jiwa) 2016-2020 [Online]. Avaiable: https://www.bps.go.id/indicator/12/1886/1/jumlah-penduduk-hasil proyeksi-menurut-provinsi-dan-jenis-kelamin.html.
[2] Nur Ika Setyani, "Perancangan Aplikasi Prediksi Harga Rumah Di Daerah Tebet Menggunakan Algoritma Regresi Linear Berbasis Machine Learning", Skripsi, Universitas Gunadarma, Depok, Indonesia, 2021.
[3] Victor Marudut Mulia Siregar, "Perancangan Aplikasi Data Mining Untuk Memprediksi Penjualan Menggunakan Metode Decision Tree Pada Apotik Ths Pematangsiantar", J. Murni Politeknik Bisnis Indonesia, vol. 7, no. 1, pp. 51-61, 2017. ISSN: 2338-8196.

[4] Indira Lutfiana Mulyahati, "Implementasi Machine Learning Prediksi Harga Sewa Apartemen Menggunakan Algoritma Random Forest Melalui Framework Website Flask Python", Tugas Akhir, Universitas Islam Indonesia, Yogyakarta, Indonesia, 2020.
[5] Luluk Suryani dan Kasmi Edy, "Application Development "Lost & Found" Android Based Using Term Frequency - Inverse Document Frequency (Tf-Idf) And Cosine Similarity Method", J. Electro Luceat, vol. 6, no. 2, Mei, pp. 1-15, 2020.
[6] Nur Fajri Azhar, "Memprediksi Waktu Memperbaiki Bug Dari Laporan Bug Menggunakan Klasifikasi Hutan Acak", Tesis, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, 2017.
[7] Probst, P., Wright, M. N., dan Boulesteix, A.-L, "Hyperparameters and Tuning Strategies for Random Forest", Application Artif Intell, London, 2018. doi:10.1002/widm.1301.
[8] Feri Irawan, Sumijan dan Yuhandri, "Prediksi Tingkat Produksi Buah Kelapa Sawit dengan Metode Single Moving Average", J. Informasi dan Teknologi, vol. 3, no.4, Juli, pp. 251-256, 2021. ISSN: 2714-9730.
[9] A. J. Syahid dan D. Mahdiana, "Perbandingan Algoritma Untuk Klasifikasi Analisis Sentimen Terhadap GeNose Pada Media Sosial Twitter", SemanTIK, vol. 7, no. 1, pp. 9-16, 2021. https://doi.org/10.5281/zenodo.5034916.
[10] Heri Setyawan, Sri Hariyati Fitriasih dan Retno Tri Vulandari, "Prediksi Tingkat Produksi Buah Kelapa Sawit dengan Metode Single Moving Average", J.TIKomSiN, vol. 9, no. 2, pp. 1-10, 2021. ISSN: 2338 4018 https://doi.org/10.30646/tikomsin.v9i2.53.

No Plagiarism Found