

**PREDIKSI PENJUALAN HARGA PROPERTI MENGGUNAKAN
ALGORITMA KNN DI METLAND CILEUNGSI**

Annisa Fitri

Universitas Mercu Buana

E-mail: annisaafitri384@gmail.com

Abstrak

Forecasting/ prediksi pasar properti adalah aktivitas yang sangat diperlukan dalam properti investasi (Mitchell dan McNamara, 1997). Dengan begitu, bisa diperkirakan berapa jumlah kebutuha yang Akan perlukan bulan depan. Peramalan produk adalah salah satu hal terpenting dalam penjualan. Dengan mengetahui prediksi harga untuk penjualan properti serta luas tanah dan bangunan . Adapun tujuan penelitian yang ingin dicapai dalam penelitian ini adalah mengetahui prediksi yang akan dilakukan oleh perusahaan sebagai strategi marketing dan mendapatkan prediksi penjualan properti berdasarkan data penjualan di tahun sebelumnya..Data mining merupakan suatu teknik untuk mengekstraksi informasi yang berharga dari data. Salah satu algoritma yang digunakan algoritma K-Nearest Neighbors. Algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan objek membutuhkan jarak terdekat dengan objek tersebut. Algoritma K-Nearest Neighbors berupaya mengklasifikasikan data baru yang belum diketahui kelasnya dengan memilih kumpulan data yang paling dekat dengan data baru tersebut. Tujuan penelitian yang ingin dicapai dalam penelitian ini adalah untuk mengetahui peramalan yang dibuat perusahaan sebagai strategi pemasaran dan memperoleh ramalan penjualan real estate berdasarkan data penjualan tahun sebelumnya. Hasil penelitian inimenunjukkan bahwa algoritma KNN dapat digunakan untuk memprediksi dengan nilai accuracy sebesar 99,8%.

Kata Kunci: *Prediksi, Data Mining, K-Nearest Neighbors.*

1. PENDAHULUAN

Pertumbuhan penduduk dan pertumbuhan ekonomi yang maju berarti kebutuhan perumahan atau tempat tinggal meningkat. Perumahan adalah salah satu bisnis property untuk investasi. Dalam bisnis ini diperlukan perhatian kebutuhan dan keinginan dari masyarakat atau konsumen. Penjualan merupakan transaksi penjual dan pembeli. Proses penjualan antara penjual dan pembeli melakukan transaksi harus di report. Penjualan ditentukan oleh harga pasar dan tingkat minat pembeli. Perusahaan real estate biasanya memasarkan beberapa jenis atau model properti yang berbeda untuk ditawarkan dengan harga yang berbeda-beda, namun tidak semua jenis tersebut memiliki jumlah peminat yang sama. Nilai real estat tidak pernah rendah, sehingga keberhasilan penjualannya sangat penting bagi keberlanjutan dan nilai perusahaan itu sendiri. Oleh karena itu, mengetahui nilai jual di masa depan dapat membantu pengembang mengidentifikasi jenis properti mana yang lebih menarik minat dan mana yang tidak. Dengan ini dapat

menggunakan peramalan / prediksi untuk memperkirakan angka penjualan. Peramalan adalah memperkirakan apa yang akan terjadi di masa depan dengan melihat situasi di masa lalu.

Metland merupakan perusahaan yang bergerak dalam bidang property. Perusahaan ini menjual berbagai type rumah, Ruko siap huni dan Kavling. Berdasarkan survey hasil minat konsumen akan produk rumah, dan ruko berdasarkan penjualan dari tahun 2013-2018 maka dibutuhkan untuk prediksi penjualan. Guna untuk mempermudah perusahaan dan team marketing untuk merekomendasikan range harga, luas rumah, yang banyak terjual dari data sebelumnya dan tim perencanaan / perusahaan bisa mendesain rumah dengan kemiripan dan harga yang lebih terjangkau. Untuk memprediksi dalam penelitian ini menggunakan data mining dengan algoritma KNN (K-Nearest Neighbor). Penelitian yang akan dilakukan dengan judul prediksi Harga penjualan properti menggunakan algoritma KNN. Tujuan penelitian ini adalah mengetahui prediksi yang akan dilakukan oleh perusahaan sebagai strategi marketing, mendapatkan prediksi penjualan properti berdasarkan data penjualan di tahun sebelumnya.

2. METODE PENELITIAN

Metode penelitian pada penelitian ini adalah metode kuantitatif. Metode analisis data kuantitatif merupakan metode yang digunakan ketika melakukan penelitian terhadap data numerik. Jenis metode ini memerlukan data numerik dalam jumlah besar, yang dapat dihitung menggunakan rumus statistik. Penggunaan metode analisis data kuantitatif merupakan hal yang banyak digunakan dalam penelitian ilmiah dan teknis.

Tahapan penelitian



Gambar 1 1 Tahapan Penelitian

a. Persiapan.

Tahap ini merupakan tahapan subjek dalam Metland Cileungsi Batasan dan menyusun rencana penelitian.

b. Studi Literatur

Sebagai bagian dari studi literatur, kami melakukan studi literatur dan mempelajari perkiraan penjualan menggunakan data terkait penjualan. Berbagai kegiatan yang berkaitan dengan cara mengumpulkan data perpustakaan, cara membaca dan mencatat, cara mengelola bahan penelitian, dan lain-lain.

c. Pengumpulan Data

Pengumpulan data dilakukan melalui wawancara dengan bagian penjualan, manajer periklanan dan pemasaran, observasi, dan dokumentasi.

Ada tiga metode pengumpulan data yang digunakan, yaitu:

- Metode observasi. Dalam hal ini data primer ditampilkan dan diperiksa secara langsung menggunakan sistem penjualan Metland Cielungsi.
- Metode Wawancara Metode ini melibatkan diskusi dan tanya jawab terhadap narasumber yang diyakini memiliki pengetahuan lebih besar tentang masalah penelitian.
- Metode Penelitian Kepustakaan Metode yang digunakan adalah dengan mencari bahan-bahan yang mendukung definisi masalah melalui buku-buku dan internet yang erat kaitannya dengan pokok permasalahan.

d. Pengolahan Data Mining

Pengolahan data mining ini data yang diolah dengan mendefinisikan KDD sebagai proses penggunaan teknik data mining untuk menemukan informasi dan pola berharga dalam data. Ini termasuk algoritma untuk mengidentifikasi pola dalam data. Meliputi pemilihan data, prapemrosesan, transformasi, penambangan data / Data Mining, dan evaluasi. Memilih data Memilih proses atau memilih data yang mungkin relevan dengan analisis. Dengan data yang dipilih, data ini dikompilasi menjadi file yang diproses untuk penambangan data. Data pemakaian merupakan data penjualan tahun 2013 sampai dengan tahun 2018. Luas tanah dan bangunan, harga, jumlah kamar tidur, dan jumlah kamar mandi dipilih. Preprocessing Proses preprocessing dari proses ini menggabungkan kumpulan data yang berbeda dan melakukan proses pembersihan. Pilih data yang ingin Anda gunakan. Transformasi Data Proses mengubah data terpilih ke dalam format untuk proses pengumpulan data. Fase ini memungkinkan Anda mengelompokkan semua data Anda berdasarkan atribut yang akan digunakan dalam proses transformasi data mining. Data Mining Metode yang digunakan dalam penelitian ini adalah metode KNN (K-Nearest Neighbors). Untuk menggunakan algoritma k-nearest neighbour, Anda perlu menentukan jumlah k-nearest neighbour yang akan digunakan untuk mengklasifikasikan data baru. Bilangan k harus bilangan ganjil, misal $k = 1, 2,$ atau 3 . Penentuan nilai k didasarkan pada banyaknya data yang ada dan besar kecilnya dimensi yang dibentuk oleh data tersebut. Semakin banyak data yang Anda miliki, semakin rendah angka k yang harus Anda pilih. Namun, semakin besar dimensi data Anda, semakin tinggi angka k yang harus Anda pilih. pengembangan pola Proses mengidentifikasi pola berdasarkan pengukuran yang diberikan. Pada tahap ini dipertimbangkan apakah informasi yang dimasukkan tidak sesuai dengan fakta atau asumsi yang ada sebelumnya. Pola penjualan real estate ini memiliki pola informasi yang dihasilkan menggunakan aturan yang berasal dari perhitungan K-Nearest Neighbor.

3. HASIL DAN PEMBAHASAN

Dataset

Dataset yang di gunakan merupakan dataset penjualan dari Metland Cielungsi dari tahun 201-2020. Mengenai Variabel yang di gunakan berisikan harga rumah, luas rumah, jumlah kamar , jumlah kamar mandi. Data yang digunakan sebanyak 1010 data dan 4 kolom.

Table 1 dataset

harga	luas	kamar 1	km	Time
1200000000	7245	3	3	01/01/2012
930000000	5934	3	3	01/03/2012
760000000	4680	2	2	01/03/2012
755000000	4680	2	2	01/08/2012
760000000	4680	2	2	01/08/2012
760000000	4680	2	2	15/01/12
905000000	5460	3	3	16/01/12
755000000	4500	2	2	24/01/12
755000000	4500	2	2	25/01/12
1300000000	9100	3	3	29/01/12
1250000000	8280	3	3	02/03/2012
760000000	4770	2	2	02/05/2012
1250000000	8280	3	3	02/06/2012
975000000	4680	2	2	02/06/2012
700000000	3744	2	2	02/12/2012
700000000	3744	2	2	02/12/2012
700000000	3744	2	2	19/02/12
700000000	2444	2	2	20/02/12
1200000000	7245	2	2	23/02/12
930000000	6195	2	2	26/02/12

```
harga luas kamar 1 km year
0 1200000000 7245 3 3 2012
1 930000000 5934 3 3 2012
2 760000000 4680 2 2 2012
3 755000000 4680 2 2 2012
4 760000000 4680 2 2 2012
... ..
1005 900000000 5032 3 3 2019
1006 400000000 3240 3 3 2019
1007 400000000 3240 3 3 2019
1008 1900000000 3744 3 3 2019
1009 1050000000 5220 3 3 2019

[1010 rows x 5 columns]
```

gambar 1 Dataset pada program python

preprocessing pada proses ini dilakukan penggabungan dataset yang berbeda dan dilakukan proses cleaning. Data dipilih data mana yang akan dipakai.

Out[7]:

	harga	luas	kamar 1	km
count	1.010000e+03	1010.000000	1010.000000	1010.000000
mean	5.118675e+09	5121.116832	2.699010	2.699010
std	6.817753e+09	3955.301555	0.458916	0.458916
min	2.576698e+08	2444.000000	2.000000	2.000000
25%	9.279091e+08	3744.000000	2.000000	2.000000
50%	2.875000e+09	4680.000000	3.000000	3.000000
75%	6.100000e+09	5220.000000	3.000000	3.000000
max	5.500000e+10	52220.000000	3.000000	3.000000

gambar 2 proses cleaning data

```
In [8]: print (data.dtypes)
harga      int64
luas       int64
kamar 1    int64
km         int64
dtype: object
```

gambar 3 klasifikasi data

Data Transformation Proses transformasi data terpilih ke dalam bentuk data mining prosedur. Pada tahap ini seluruh data dapat dikelompokkan dengan atribut yang digunakan dalam proses transformasi data mining pada gambar 4.10 Membagi dataset sintesis menjadi set pelatihan dan pengujian, peneliti dapat melatih model dengan menggunakan subset pelatihan (X_train dan y_train) dan menguji performa model dengan menggunakan subset pengujian (X_test dan y_test) terlihat pada gambar .

```
In [9]: features = ['harga', 'luas', 'kamar 1', 'km']
x = data[features]
y = data['harga']
x.shape, y.shape
```

Out[9]: ((1010, 4), (1010,))

gambar 4 transformasi data

Split data dilakukan untuk menguji performa model yang dikembangkan dengan menggunakan subset pelatihan pada subset pengujian yang independen. Dengan memisahkan dataset menjadi dua subset yang berbeda, peneliti dapat menguji sejauh mana model dapat menggeneralisasi pola dari subset pelatihan ke subset pengujian yang belum pernah dilihat sebelumnya. Dalam penelitian ini mebagi data 70 – 30.

```
In [10]: from sklearn.model_selection import train_test_split
train_x, test_x, train_y, test_y = train_test_split(x, y, random_state = 70)
```

```
In [11]: test_y.shape
```

Out[11]: (253,)

gambar 5 split data

Pembuatan Model

Tahapan selanjutnya peneliti melakukan Transform Data, Transform Data adalah proses mengubah atau memanipulasi data dari satu bentuk ke bentuk lain dengan tujuan untuk memperbaiki karakteristik data, memenuhi asumsi analisis, atau meningkatkan performa model. Transformasi data dapat dilakukan pada berbagai jenis data, termasuk

data numerik dan kategorikal. Transformasi data sering dilakukan sebagai bagian dari pemrosesan data sebelum analisis statistik atau pembuatan.

Data Mining Metode yang digunakan pada penelitian ini adalah metode KNN (K-Nearest Neighbors). Untuk menggunakan algoritma k-nearest neighbour, perlu menentukan jumlah k-nearest neighbour yang akan digunakan untuk mengklasifikasikan data baru. Dalam penelitian ini menggunakan model algoritma KNN dengan menggunakan analisis jarak terdekat. Untuk modeling pertama menggunakan parameter = 5. Pegambilan prediksi menggunakan 5 tetangga terdekat.

```
In [15]: from sklearn.neighbors import KNeighborsRegressor as KNN_Reg
from sklearn.metrics import mean_squared_error as mse

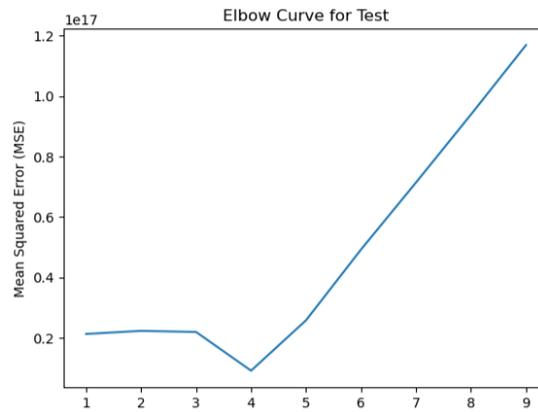
model = KNN_Reg(n_neighbors = 5)

# training the model:
model.fit(train_x, train_y)
acc1 = model.score(test_x, test_y)

# test for prediction
test_predict = model.predict(test_x)
score = mse(test_predict, test_y)
print(' MSE: ', score, '\n', 'Accuracy: ', acc1)

MSE: 2.5043779375759156e+16
Accuracy: 0.9993591158288321
```

gambar 6 pembuatan model knn



gambar 7 visualisasi grafik data

Pengujian Data

Validasi data adalah proses mengidentifikasi pola berdasarkan pengukuran yang diberikan. Fase ini juga menguji apakah informasi yang terkandung sesuai dengan fakta dan asumsi yang telah ada sebelumnya. Pola dalam penjualan property ini memiliki pola informasi yang dihasilkan adalah rules yang di dapat dari perhitungan K-Nearest Neighbours. Pada gambar 4.9 ada 2 tahap pengujian yang pertama model old dengan parameter k = 5 , dan new_model dengan parameter k =7 dan prediction hasil accuracy new model dengan nilai k = 7 mendapat acuracy 99.8% sedangkan hasil dari accuracy old model dengan parameter k = 5 mendapat accuracy 99,9% improvment 0.1%.

```
In [25]: new_model = KNN_Reg(n_neighbors=7)

# Train model
new_model.fit(train_x, train_y)
acc2 = new_model.score(test_x, test_y)

# Prediction test
print(' Accuracy of new model (%):', acc2*100, '\n', 'Accuracy of old model (%):', acc1*100, '\n Improvement (%):', (acc2-acc1)*100)

Accuracy of new model (%): 99.81643962225604
Accuracy of old model (%): 99.93377453515238
Improvement (%): -0.11733491289633857
```

gambar 8 pengujian data

Analisis Hasil

Berdasarkan penelitian menggunakan metode k nearest neighbours dengan split data 0,3 dan 0,7 dengan random state 70. Hasil prediksi dari setiap model disimpan dalam variabel 'prediction old dan prediction new', dan kemudian dilakukan penghitungan akurasi menggunakan fungsi 'acc1 untuk nilai accuracy prediction old '. Nilai akurasi ini merupakan proporsi dari jumlah prediksi yang benar dibandingkan dengan jumlah total data pengujian. peneliti dapat memperoleh hasil akurasi dari model KNN dengan berbagai jumlah tetangga (k) yang berbeda. Hal ini berguna untuk mengevaluasi performa model KNN dengan variasi nilai k dan memilih nilai k yang memberikan akurasi terbaik.

4. KESIMPULAN

Berdasarkan hasil pengujian yang dilakukan oleh peneliti dengan bantuan tools python, dapat disimpulkan bahwa penelitian ini berhasil mengetahui tingkat performa algoritma K-Nearest Neighbours. Hasil pengujian accuracy prediksi menggunakan metode KNN dengan split data 0,3 dan 0,7 dengan nilai K = 5 menghasilkan accuracy 99,9% dan pengujian nilai k =7 menghasilkan accuracy 99,8% dengan nilai improvement 0,1%. Dalam penelitian ini disimpulkan bahwa semakin tinggi nilai K maka hasil accuracy semakin rendah.

DAFTAR PUSTAKA

- [1] M. Kafil and F. T. Industri, "Penerapan Metode K-Nearest Neighbors," vol. 3, no. 2, pp. 59–66, 2019.
- [2] F. Hermawan and H. Agung, "Implementasi Metode K-Nearest Neighbor Pada Aplikasi Data Penjualan PT. Multitek Mitra Sehati," J. Sains dan Teknol., vol. 4, pp. 103–109, 2017, [Online]. Available: <http://research.kalbis.ac.id/Research/Files/Article/Full/CDRHHLKPOJ7STTQHS1YJF02JX.pdf>
- [3] H. S. Depan, "IMPLEMENTASI METODE K-NEAREST NEIGHBOR," 2021.
- [4] A. Ghani Muttaqin, K. Auliasari, and F. Santi Wahyuni, "Penerapan Metode K-Nearest Neighbor Untuk Prediksi Penjualan Berbasis Web Pada Pt.Wika Industri Energy," JATI (Jurnal Mhs. Tek. Inform., vol. 4, no. 2, pp. 1–6, 2020, doi: 10.36040/jati.v4i2.2728.
- [5] M. M. K. Neighbor, "Penerapan data mining untuk prediksi penjualan produk elektronik terlaris menggunakan metode k-nearest neighbor," 2018.
- [6] R. N. Sukmana, Abdurrahman, and Y. Wicaksono, "Implementasi K-Nearest Neighbor Untuk Menentukan Prediksi Penjualan (Studi Kasus : PT Maksipus Utama Indonesia)," J. Teknol. Inf. dan Komun. Vol. 8 No. 2, Desember 2020, vol. 8, no. 2, pp. 31–38, 2020.
- [7] A. Kataria and M. D. Singh, "A Review of Data Classification Using K-Nearest Neighbour Algorithm," Int. J. Emerg. Technol. Adv. Eng., vol. 3, no. 6, pp. 354–360, 2013.
- [8] S. Wiyono and T. Abidin, "Implementation of K-Nearest Neighbour (Knn) Algorithm To Predict Student'S Performance," Simetris J. Tek. Mesin, Elektro dan Ilmu Komput., vol. 9, no. 2, pp. 873–878, 2018, doi: 10.24176/simet.v9i2.2424.
- [9] S. C. Tékouabou Koumético and H. Touluni, "Improving KNN Model for Direct Marketing Prediction in Smart Cities," Stud. Comput. Intell., vol. 971, no. July, pp. 107–118, 2021, doi: 10.1007/978-3-030-72065-0_7.
- [10] P. Studi, T. Informatika, F. I. Komputer, and U. M. Buana, "Implementasi K-Means & Google Data Studio Untuk Clustering Produktivitas Pertanian Pada 10 Kabupaten / Kota di Jawa Barat Implementasi K-Means & Google Data Studio Untuk Clustering Produktivitas Pertanian Pada 10 Kabupaten / Kota di Jawa Barat," 2022.
- [11] I. D. Oktaviani and A. G. Putrada, "KNN imputation to missing values of regression-based

- rain duration prediction on BMKG data,” J. Infotel, vol. 14, no. 4, pp. 249–254, 2022, doi: 10.20895/infotel.v14i4.840.
- [12] T. M. Kadarina and M. H. Ibnu Fajar, “Pengenalan Bahasa Pemrograman Python Menggunakan Aplikasi Games Untuk Siswa/I Di Wilayah Kembangan Utara,” J. Abdi Masy., vol. 5, no. 1, p. 11, 2019, doi: 10.22441/jam.2019.v5.i1.003.