

SKRIPSI

**ANALISIS MODEL PREDIKSI
HARGA JUAL RUMAH DI KOTA BANDUNG
MENGUNAKAN METODE POHON REGRESI KERNEL**



ARIELLA VANIA LYNN

NPM: 6161801012

**PROGRAM STUDI MATEMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2022**

FINAL PROJECT

**ANALYSIS OF HOUSE SELLING PRICE
PREDICTION MODEL IN BANDUNG
USING KERNEL REGRESSION TREE METHOD**



ARIELLA VANIA LYNN

NPM: 6161801012

**DEPARTMENT OF MATHEMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2022**

LEMBAR PENGESAHAN

ANALISIS MODEL PREDIKSI
HARGA JUAL RUMAH DI KOTA BANDUNG
MENGUNAKAN METODE POHON REGRESI KERNEL

ARIELLA VANIA LYNN

NPM: 6161801012

Bandung, 8 Juli 2022

Menyetujui,

Pembimbing 1



Agus Sukmana, M.Sc.

Pembimbing 2



Rizky Reza Fauzi, D.Phil.Math.

Ketua Tim Penguji



Iwan Sugiarto, M.Si.

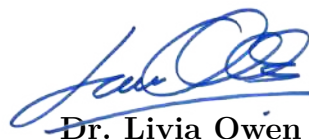
Anggota Tim Penguji



Taufik Limansyah, M.T.

Mengetahui,

Ketua Program Studi



Dr. Livia Owen

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

ANALISIS MODEL PREDIKSI HARGA JUAL RUMAH DI KOTA BANDUNG MENGUNAKAN METODE POHON REGRESI KERNEL

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
Tanggal 8 Juli 2022



Ariella Vania Lynn
NPM: 6161801012

ABSTRAK

Rumah yang merupakan salah satu contoh dari properti, dapat diperjualbelikan atau disewakan. Namun, dalam menetapkan harganya terdapat beberapa kesalahan yang sering dilakukan pihak bersangkutan, yaitu mengutamakan kepentingan pribadi, tidak memiliki cukup informasi, dan tidak menyeimbangkannya dengan kondisi bangunan. Salah satu metode statistika yang dapat digunakan untuk membantu memprediksi harga jual atau sewa sebuah rumah adalah pohon regresi. Pohon regresi merupakan model yang berbentuk hierarki dan berfungsi untuk memprediksi variabel prediktor data jenis numerik kontinu melalui pemanfaatan nilai rata-rata (mean), dengan cara kerja mengelompokkan. Namun dengan sifat tersebut, meskipun model menjadi mudah untuk dipahami, hasil prediksi yang diberikan umumnya hanya memiliki akurasi sekitar 40% sampai 60%, sehingga diperlukan metode lain untuk dikombinasikan atau sebagai pengembangan agar membantu kinerja. Salah satu metode yang dapat dikombinasikan dengan pohon regresi adalah regresi kernel. Regresi kernel merupakan model yang berfungsi untuk memprediksi variabel prediktor data numerik melalui pemanfaatan fungsi kepadatan peluang. Dengan sifat tersebut, hasil prediksi yang diberikan model dapat mencapai akurasi sekitar 70% sampai 90%, karena nilai yang ditunjukkannya sangat rinci. Pada skripsi ini, dibahas model pohon regresi kernel untuk memprediksi harga jual rumah di Kota Bandung. Model tersebut dibangun dengan melibatkan algoritma pembelajaran mesin agar hasil yang diberikan dapat dievaluasi dan dikembangkan secara optimal. Model pohon regresi kernel yang dihasilkan tersebut memiliki akurasi sebesar 78%, sehingga sudah dapat diterapkan pada data penjualan rumah di Kota Bandung dengan baik.

Kata-kata kunci: harga jual rumah, pembelajaran mesin, pohon regresi, regresi kernel, pohon regresi kernel.

ABSTRACT

House as an example of a property, can be sold or rented. However, in determining its price, there exists several mistakes that are often made by the parties related, such as prioritizing personal interests, not having enough information, and not balancing it with the building condition. One of the statistical methods that can be used to help predict the selling or rental price of a house is regression tree. Regression tree is a hierarchical model and functions to predict continuous numeric type of predictor variable data by grouping through the use of average value (mean). However, with these traits, even though the model is easy to understand, the prediction results given are generally only has accuracy around 40% to 60%, so other methods are needed to be combined or as developments to help the performance. One method that can be combined with regression tree is kernel regression. Kernel regression is a model that functions to predict numeric type of predictor variable data through the use of probability density function (PDF). Now, with these traits, the prediction results given by the model can be reach accuracy around 70% to 90%, because the value shown is pretty detailed. In this undergraduate thesis, a kernel regression tree model to predict the selling price of houses in Bandung is discussed. The model was built by involving machine learning algorithms, so that the given results can be evaluated and developed optimally. The results of this kernel regression tree model has accuracy of 78%, so it can be applied to data of house sales in Bandung properly.

Keywords: house selling price, machine learning, regression tree, kernel regression, kernel regression tree.

Untuk papa, mama, dan adik tersayang

KATA PENGANTAR

Puji dan syukur kepada Tuhan dan Bunda Maria atas anugerah-Nya, penulis mampu mengerjakan skripsi ini dengan baik. Skripsi yang berjudul "Analisis Model Prediksi Harga Jual Rumah di Kota Bandung" disusun sebagai salah satu syarat untuk menyelesaikan studi Strata-1 Program Studi Matematika, Fakultas Teknologi Informasi dan Sains, Universitas Katolik Parahyangan, Bandung.

Selama pengerjaan skripsi ini, penulis memperoleh banyak ilmu dan pertolongan dari berbagai pihak, sehingga mampu menghadapi beragam tantangan. Oleh karena itu, penulis mengucapkan terima kasih terkhususnya kepada:

1. Papa, mama, dan adik yang selalu mendoakan, mendukung, dan menghibur penulis sepanjang bergumul dengan skripsi ini.
2. Bapak Dr. Putu Harry Gunawan selaku satu dosen tamu untuk Mata Kuliah Pilihan Kapita Selektiva Statistika tahun ajaran 2020-2021 yang telah memberikan referensi contoh data dan pemrograman mengenai sains data, sehingga penulis menjadi terinspirasi untuk mengerjakan skripsi ini.
3. Bapak Agus Sukmana, M.Sc. selaku dosen pembimbing 1 dan Bapak Rizky Reza Fauzi, D.Phil.Math. selaku dosen pembimbing 2 yang dengan sabar telah memberikan ilmu dan memaklumi kemampuan penulis, serta bersedia membagikan pengalaman dan meluangkan waktu di luar jadwal kuliah, sehingga penulis menjadi lebih termotivasi dalam mengerjakan skripsi ini.
4. Bapak Iwan Sugiarto, M.Si. selaku dosen penguji 1 dan Bapak Taufik Limansyah, S.Si., M.T. selaku dosen penguji 2 yang telah memberikan kritik dan saran agar penulis mampu menciptakan skripsi ini dengan lebih berkualitas.
5. Bapak Dr. Daniel Salim selaku koordinator skripsi yang telah memberikan arahan demi terselesaikannya skripsi ini.
6. Seluruh dosen Universitas Katolik Parahyangan, terutama dari Program Studi Matematika yang telah memberikan ilmu dan didikan.
7. Kakak Jennifer Lorenza yang telah bersedia meluangkan waktu untuk berbagi ilmu dan pengalaman sepanjang penulis berkuliah, terutama dalam mengerjakan skripsi ini.
8. Felicia Iona Thamara, Vianca Nethania Gunawan, Kaschia Sherina, Feliya, Ambrosius Adrian Djunaidi, dan Elbert yang telah bersedia menjadi tutor sepanjang penulis berkuliah.
9. Meilyana Novita Silaban yang telah menemani penulis sepanjang berkuliah secara luring, serta Anthony Jason Susilo dan Rhandy Ghetly Wirawan yang telah menemani penulis sepanjang berkuliah secara daring.

Oleh karena keterbatasan kemampuan dan pengalaman, penulis menyadari bahwa skripsi ini masih jauh dari sempurna, sehingga akan dengan senang hati menerima kritik dan saran yang membangun dari para pembaca. Penulis berharap skripsi ini dapat memberikan manfaat bagi semua pembacanya.

Bandung, Juli 2022

Penulis

DAFTAR ISI

KATA PENGANTAR	xv
DAFTAR ISI	xvii
DAFTAR GAMBAR	xix
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	2
1.4 Batasan Masalah	2
1.5 Metodologi	3
1.6 Sistematika Pembahasan	3
2 LANDASAN TEORI	5
2.1 Model	5
2.2 Analisis Prediktif	5
2.3 Model Regresi	6
2.4 Pembelajaran Mesin	6
2.5 Penambangan Data	8
2.6 Pohon Keputusan	8
2.6.1 Visualisasi Pohon Regresi	9
2.6.2 Pembentukan Pohon Regresi	11
2.6.3 Interpretasi Pohon Regresi	14
2.6.4 Evaluasi Pohon Regresi	15
2.6.5 Keunggulan dan Kelemahan Pohon Regresi	17
2.7 Estimasi Kepadatan Kernel	17
2.8 Pohon Regresi Kernel	21
3 METODE	23
3.1 Data	23
3.2 Prosedur	24
4 HASIL DAN PEMBAHASAN	27
4.1 Persiapan Data	27
4.2 Model Pohon Regresi I	27
4.3 Model Pohon Regresi II	27
4.3.1 Akar	28
4.3.2 Node Internal 1	28
4.3.3 Daun 1	29
4.3.4 Node Internal 2	29
4.3.5 Daun 2	29
4.3.6 Node Internal 3	29

4.3.7	Node Internal 4	30
4.3.8	Node Internal 5	30
4.3.9	Daun 3	30
4.3.10	Daun 4	30
4.3.11	Daun 5	30
4.3.12	Node Internal 6	31
4.3.13	Daun 6	31
4.3.14	Node Internal 7	31
4.3.15	Node Internal 8	31
4.3.16	Daun 7	32
4.3.17	Daun 8	32
4.3.18	Daun 9	32
4.3.19	Daun 10	32
4.3.20	Evaluasi	33
4.4	Statistika Deskriptif Data Kategorik	33
4.5	Model Pohon Regresi Kernel	33
4.5.1	Model Regresi Kernel untuk Pohon Regresi II Daun 2	41
4.5.2	Model Regresi Kernel untuk Pohon Regresi II Daun 4	42
4.5.3	Model Regresi Kernel untuk Pohon Regresi II Daun 6	43
4.5.4	Model Regresi Kernel untuk Pohon Regresi II Daun 7	44
4.5.5	Model Regresi Kernel untuk Pohon Regresi II Daun 8	45
4.5.6	Model Regresi Kernel untuk Pohon Regresi II Daun 10	46
4.5.7	Evaluasi	46
4.6	Penerapan	47
5	PENUTUP	49
5.1	Kesimpulan	49
5.2	Saran	50
	DAFTAR REFERENSI	51

DAFTAR GAMBAR

2.1 Tahapan Analisis Prediktif	5
2.2 Tahapan Pembelajaran Mesin	7
2.3 Tahapan Pembelajaran Mesin Terawasi	7
2.4 Tahapan Penambangan Data	8
2.5 Contoh Struktur Pohon Keputusan	9
2.6 Visualisasi I Pohon Regresi 1 Variabel Prediktor	9
2.7 Visualisasi II Pohon Regresi 1 Variabel Prediktor	9
2.8 Visualisasi I Pohon Regresi 2 Variabel Prediktor	10
2.9 Visualisasi II Pohon Regresi 2 Variabel Prediktor	10
2.10 Persebaran Contoh Data	11
2.11 Percobaan I Membagi Daerah Persebaran Data	12
2.12 Galat Seluruh Percobaan Membagi Daerah Data	12
2.13 Pembatas Terbaik Persebaran Data	13
2.14 Percobaan I Gelombang II Membagi Daerah Persebaran Data	13
2.15 Pembatas Terbaik II Persebaran Data	13
2.16 Model Hampiran Pohon Regresi Terhadap Contoh Data	14
2.17 Model Pohon Regresi Contoh Data	14
2.18 Model Pohon Regresi Contoh Data Menggunakan Python	15
2.19 Contoh Visualisasi Uji <i>Underfitting</i> dan <i>Overfitting</i>	16
2.20 Bentuk Hampiran Model	16
2.21 Contoh Kurva Hampiran Kernel Gaussian	18
2.22 Contoh Hampiran Data oleh KDE dengan Beragam Lebar Pita	19
2.23 Contoh Model Regresi Kernel Jenis Gaussian	21
2.24 Contoh Perbandingan antara Grafik Pohon Regresi dan Regresi Kernel	22
4.1 Uji <i>Underfitting</i> dan <i>Overfitting</i> Data	28
4.2 Model Pohon Regresi Data	34
4.3 Keadaan dari Variabel Prediktor Kecamatan dan Kompleks	35
4.4 Keadaan dari Variabel Prediktor Bagian Dalam	36
4.5 Keadaan dari Variabel Prediktor Sumber Air	36
4.6 Keadaan dari Variabel Prediktor Garasi	36

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Properti merupakan aset yang memiliki harga pasar selalu meningkat setiap tahunnya. Menurut salah satu pengembang real estat di Indonesia, PT Pudjiadi Prestige Tbk ¹, hal tersebut terjadi karena adanya efek infrastruktur, serta penambahan jumlah penduduk dan kebutuhan akan hunian yang terus bertambah, sedangkan porsi ketersediaan tempat itu sendiri tidak kunjung bertambah. Perkembangan infrastruktur pada suatu daerah memengaruhi harga sebuah properti, di mana jika terdapat penambahan fasilitas tertentu, harga properti di daerah tersebut akan menjadi naik, dan jika terdapat pengurangan fasilitas, harga properti menjadi turun. Kemudian, penambahan jumlah penduduk yang tidak disertai dengan penambahan tempat hunian mengakibatkan setiap orang berlomba untuk mendapatkan tempat hunian tertentu atau menjualnya dengan harga yang tinggi.

Salah satu contoh dari properti adalah rumah, yang pada umumnya memiliki empat bagian utama, yaitu ruang keluarga, dapur, kamar tidur, dan kamar mandi. Kemudian, aspek yang menjadi perhatian pada sebuah rumah adalah keamanan, kenyamanan, serta kedekatan dan kemudahan untuk mengakses fasilitas umum. Sifat yang tergolong sebagai properti, membuat rumah dapat diperjualbelikan atau disewakan. Hal yang perlu dipersiapkan untuk menjual atau menyewakan sebuah rumah terdiri atas dokumen resmi seperti rekening sumber daya dan spesifikasi berbagai unsur seperti perincian keadaan dalam dan luar rumah. Seorang penjual atau penyewa sebuah rumah tidak dapat asal menetapkan harga, tetapi harus terlebih dahulu melakukan analisis terhadap harga pasar dan berbagai faktor yang dimiliki oleh rumah bersangkutan [1].

Berdasarkan catatan Kompas ², kesalahan yang sering terjadi saat menetapkan harga jual (sewa) sebuah rumah adalah mengutamakan kepentingan pribadi, memasuki pasar properti tanpa memiliki cukup informasi, dan tidak menyeimbangkannya dengan kondisi bangunan. Seseorang yang menetapkan harga jual (sewa) sebuah rumah dengan mengutamakan kepentingan pribadi seperti terobsesi untuk memperoleh keuntungan yang besar, membuka peluang diperolehnya keuntungan yang tidak maksimum karena harga yang ditetapkan menjadi terlalu tinggi. Kemudian, seseorang yang menetapkan harga jual (sewa) sebuah rumah tanpa memiliki cukup informasi seputar properti, menjadikan harga yang ditetapkannya tidak sesuai dengan keadaan yang tengah berlangsung. Selanjutnya, seseorang yang menetapkan harga jual atau sewa sebuah rumah dengan tidak menyeimbangkannya terhadap kondisi bangunan itu sendiri, dapat mengakibatkan calon pelanggan segera beralih kepada properti lain karena harga yang ditetapkannya tidak rasional.

Seiring berkembangnya ilmu pengetahuan dan teknologi (IPTEK), terdapat algoritma pembelajaran mesin yang dapat membantu manusia melakukan analisis suatu perkara secara lebih baik, khususnya mengenai prediksi. Hal ini karena rangkaian kinerja yang dilakukan dapat dievaluasi dan dikembangkan melalui beragam teknik, sehingga menjadi lebih optimal. Metode statistika yang

¹PT Pudjiadi Prestige. "Ini Dia Alasan Mengapa Harga Properti Terus Naik Tiap Tahun". (2020). <https://pudjiadiprestige.co.id/news/ini-dia-alasan-mengapa-harga-properti-terus-naik-tiap-tahun.html>, diakses tanggal 13 November 2021.

²Kompas. "Catat... 10 Kesalahan Ketika Menjual Rumah!". (2013). <https://properti.kompas.com/read/2013/09/03/0442095/Catat.10.Kesalahan.Ketika.Menjual.Rumah.?page=all>, diakses tanggal 16 November 2021.

umum digunakan untuk membantu memprediksi harga jual atau sewa sebuah rumah adalah regresi linear, tetapi terdapat pula metode lain yang dapat digunakan, salah satunya adalah pohon regresi [2]. Pohon regresi merupakan model yang berbentuk hierarki dan berfungsi untuk memprediksi variabel prediktor data jenis numerik kontinu melalui pemanfaatan nilai rata-rata (mean), dengan cara mengelompokkan.

Pada skripsi ini, dibahas model pohon regresi untuk memprediksi harga jual rumah di Kota Bandung, dengan proses pembuatannya melibatkan algoritma pembelajaran mesin. Metode ini dipilih karena memiliki keunggulan berupa mampu mengolah data secara sederhana tanpa memerlukan asumsi tertentu, terbuka terhadap data kategorik dan numerik dengan skala yang beragam tanpa perlu melakukan normalisasi, serta cara menginterpretasikannya mudah sehingga ramah digunakan oleh semua pihak yang membutuhkan [3]. Akan tetapi, meskipun memiliki keunggulan demikian, pohon regresi juga memiliki kelemahan, salah satunya adalah proses regresi dilakukan dengan cara pengelompokkan, di mana total data yang ada dibagi ke dalam beberapa kelompok (daerah) dan memprediksi hanya berdasarkan kelompok tersebut, sehingga tidak memberikan hasil yang rinci mengenai setiap individu data. Untuk mengatasi kelemahan tersebut, pada skripsi ini dipilih juga salah satu metode statistika lain untuk dikombinasikan dengan pohon regresi, yaitu regresi kernel. Regresi kernel memiliki karakteristik yang serupa dengan pohon regresi, hanya saja proses regresinya dilakukan dengan memanfaatkan fungsi kepadatan peluang (*Probability Density Function* - PDF) tanpa melibatkan hierarki data [4]. Melalui pohon regresi kernel ini, diharapkan masyarakat di Kota Bandung dapat menetapkan harga jual rumah dengan lebih optimal.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam skripsi ini adalah:

1. Bagaimana karakteristik model pohon regresi yang dapat digunakan untuk memprediksi harga jual rumah di Kota Bandung?
2. Bagaimana karakteristik model pohon regresi yang telah dikombinasikan dengan model regresi kernel untuk memprediksi harga jual rumah di Kota Bandung?

1.3 Tujuan

Berdasarkan rumusan masalah yang telah dipaparkan, tujuan dari skripsi ini adalah:

1. Mengetahui karakteristik model pohon regresi yang dapat digunakan untuk memprediksi harga jual rumah di Kota Bandung.
2. Mengetahui karakteristik model pohon regresi yang telah dikombinasikan dengan model regresi kernel untuk memprediksi harga jual rumah di Kota Bandung.

1.4 Batasan Masalah

Batasan masalah dalam skripsi ini adalah:

1. Data yang digunakan berasal dari salah satu situs properti, yaitu <https://www.rumah.com/>.
2. Data yang diambil terdapat sebanyak lima ratus sampel, dengan fokus wilayah beberapa kompleks perumahan di Kota Bandung.
3. Data terdiri atas tiga belas atribut yaitu Kecamatan, Kompleks, Bagian Dalam, Harga Tanah, Luas Tanah, Luas Bangunan, Sumber Air, Tingkat, Kamar Mandi, Kamar Tidur, Garasi, Daya Listrik, dan Harga Jual.
4. Data diolah menggunakan bahasa pemrograman Python, serta perangkat lunak Microsoft Excel dan Tableau.

1.5 Metodologi

Metode penelitian pada skripsi ini adalah studi kasus, di mana analisis yang dilakukan memanfaatkan proses penambangan data. Data yang digunakan diperoleh dengan mencatat secara manual dalam perangkat lunak Microsoft Excel, lalu data dirapikan agar siap diproses lebih lanjut. Setelah itu, model yang diperlukan dibangun menggunakan bahasa pemrograman Python dan statistika deskriptif data dipaparkan menggunakan perangkat lunak Tableau, lalu mengevaluasi hasilnya.

1.6 Sistematika Pembahasan

Skripsi ini terdiri atas lima bab, yaitu:

Bab 1: Pendahuluan

Bab ini berisi latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, metodologi, dan sistematika pembahasan skripsi.

Bab 2: Landasan Teori

Bab ini berisi teori model, analisis prediktif, analisis regresi, pembelajaran mesin, penambangan data, pohon keputusan, estimasi kepadatan kernel, dan pohon regresi kernel.

Bab 3: Metode

Bab ini berisi uraian data yang digunakan dan prosedur yang dilakukan untuk menganalisis dalam skripsi.

Bab 4: Hasil dan Pembahasan

Bab ini berisi analisis model pohon regresi kernel yang diperoleh beserta eksplorasi datanya, dengan menunjukkan hasil visualisasi dan interpretasinya.

Bab 5: Penutup

Bab ini berisi kesimpulan yang dapat ditarik berdasarkan pembahasan hasil dan saran untuk pengembangan penelitian selanjutnya.