

**ANALISIS MODEL PREDIKSI BISNIS PROPRTI PADA DATA E-COMMERCE
DENGAN METODE KLASIFIKASI**

**ANALYSIS PREDICTION MODEL OF PROPERTY BUSINESS BASED ON E-
COMMERCE DATA USING CLASSIFICATION METHODS**

Fariz Denada Sudrajat¹, Andry Alamsyah²

Prodi S1 Manajemen Bisnis Telekomunikasi dan Informatika, Universitas Telkom

Email : farizdenada@gmail.com¹, andry.alamsyah@gmail.com²

ABSTRAK

Properti bisnis online atau e-commerce yang diketahui saat ini mengalami peningkatan penjualan rumah. E-Commerce di Indonesia dari tahun 2011 sampai 2015 telah meningkat lebih dari 500%. Usaha real estat di Indonesia mengalami tren positif. Prediksi harga itu penting karena bisa bermanfaat bagi investor atau masyarakat untuk membantu setiap orang untuk membeli rumah. Salah satu metode prediksi adalah klasifikasi berdasarkan variabel industri properti.

Di era ini, data mudah didapat. Ada banyak open source dari internet yang bisa kita dapatkan secara gratis. Data e-commerce adalah salah satu data yang bisa kita gunakan untuk penelitian. Data bisa digunakan untuk semua orang untuk melakukan penelitian.

Untuk memprediksi penjualan real estat, penelitian ini menggunakan 2 metode klasifikasi yang berbeda. Peneliti memilih Bandung sebagai studi kasus berdasarkan data dari e-commerce rumah123.com. Metodenya adalah klasifikasi Decision Tree dan k-NN. Peneliti membuat analisa perbandingan antara Decision Tree dan k-NN. Hasilnya akurasi pohon keputusan adalah 75% dan akurasi hasil KNN adalah 71%.

Kata Kunci: E-Commerce, Data Mining, Klasifikasi

ABSTRACT

Online business property or known e-commerce is currently experiencing an increase of home sales. E-Commerce in Indonesia from 2011 to 2015 has increased by more than 500%. Real estates business in Indonesia are in a positive trending. Prediction of price is important because this can be useful for the investor or public to help everyone to buy home. One of methods for prediction is classification based on variables of property industry.

In this era, data is easily obtained. There are many open source from internet that we can get for free. E-commerce data is one of data that can we use for research. Data can be used for everyone to do a research.

In order to predict real estate sales, this research employed 2 different classification methods. Researchers choose Bandung as study case based on data from e-commerce rumah123.com. the methods are Decision Tree and k-NN classification. Researcher made a comparison between Decision Tree and k-NN. The result is decision tree accuracy is 75% and KNN accuracy result is 71%.

Keywords: *E-Commerce, Data Mining, Classification*

1. Latar Belakang Masalah

Pada zaman modern saat ini, Internet merupakan kebutuhan manusia dalam sehari-hari. Manfaat internet bagi masyarakat di Indonesia memang cukup banyak dan sangat membantu dalam kehidupan sehari-hari. IWS mencatat Indonesia termasuk 5 besar dikawasan Asia sebagai pengguna internet terbanyak.^[1]

Di Indonesia, nilai transaksi e-commerce dalam lima tahun terakhir mengalami pertumbuhan positif. berdasarkan data yang dihimpun oleh tim riset Tirto.id dari statista, nilai transaksi e-commerce (B2C) di Indonesia dalam periode 2011 hingga 2015 telah tumbuh lebih dari 500 persen, mencapai angka \$3,56 miliar. Seiring dengan pertumbuhan tingkat penetrasi internet masyarakat Indonesia yang diprediksi akan terus naik hingga dua kali lipat pada tahun 2020 dari posisi saat ini yakni 30 persen, maka prospek cerah industri ini menjadi semakin tidak terbantahan. Menurut riset dari Macquarie, pasar e-commerce Indonesia diperkirakan akan tumbuh hingga 80 persen menjadi \$7 miliar pada tahun 2016.^[2]

Dilansir dari okezone, Bisnis properti sangat erat kaitannya dengan dua hal, infrastruktur dan kondisi alam. Itulah yang menjadi dasar mengapa pada akhirnya wilayah Bandung Timur dan Selatan saat ini dan juga ke depannya diproyeksikan menjadi surga bagi lokasi.^[3]

Data Mining adalah proses menemukan pola wawasan, menarik, dan yang tidak terlihat, sedeskriptif mungkin, dapat dimengerti dan dapat memprediksi model dari data skala besar.^[5] Klasifikasi memetakan data menjadi kelompok-kelompok yang telah ditetapkan dari kelas. Hal ini sering disebut sebagai pembelajaran diawasi karena kelas ditentukan sebelum memeriksa data.^[7]

2. Landasan Teori

2.1 Big Data

Big Data adalah data yang melebihi kapasitas pengolahan sistem database konvensional. Data tersebut terlalu besar, bergerak terlalu cepat, atau tidak sesuai dengan arsitektur basis data. Untuk mendapatkan nilai dari data ini harus memilih *alternative* cara untuk mengelolanya untuk memprosesnya.^[11]

2.2 E-Commerce

Penggunaan internet dan Web untuk transaksi bisnis lebih formal digital memungkinkan transaksi komersial antara dan di antara organisasi dan individu.^[12]

2.3 CRISP-DM

CRISP-DM merupakan metode yang bebas untuk dipakai bagi siapa saja untuk meningkatkan keberhasilan data mining. Terdapat enam tahap dalam metode ini, dan setiap tahap berjenis adaptif yang artinya tahap selanjutnya bergantung pada hasil tahap sebelumnya.^[10]

2.4 Data Mining

Data Mining adalah proses menemukan pola wawasan, menarik, dan yang tidak terlihat, sedeskriptif mungkin, dapat dimengerti dan dapat memprediksi model dari data skala besar.^[5]

2.5 Klasifikasi

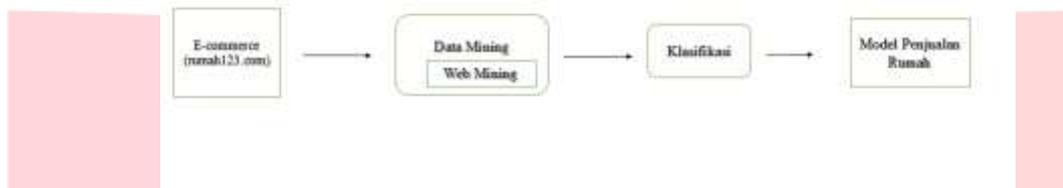
Klasifikasi ditujukan pada tugas untuk memprediksi sebuah label kelas pada poin yang belum berlabel.^[5]

2.6 Web Mining

Web data mining menggunakan teknologi data mining ke Web, dan abstrak informasi yang menarik dari teks, informasi, data dan jasa lainnya. Teknologi ini membentang lebih dari beberapa teknik komputer dengan bahasa komputer, statistik, kecerdasan buatan, teknologi jaringan dan informasi komputer.^[8]

2.7 Kerangka Pemikiran

Tahap pertama dari penelitian ini adalah melihat data e-commerce disitus rumah123, lalu mengambil data tersebut dan mengubahnya menjadi model klasifikasi penjualan rumah.



Gambar 1 Kerangka Pemikiran

Sumber: Data yang diolah

3. Metode Penelitian

3.1 Karakteristik Penelitian

Peneliti menggunakan metode Data Mining untuk melakukan Knowledge Discovery in Database (KDD), proses *non trivial* dari mengidentifikasi valid, baru, berpotensi, berguna, dan mudah dimengerti dalam data.^[14]

3.2 Alat Pengumpulan Data

Peneliti mengumpulkan data dengan cara web mining pada situs rumah123.com dengan menggunakan ParseHub.

3.3 Tahapan Penelitian

Tahapan penelitian yang peneliti lakukan yaitu mengidentifikasi masalah, lalu membuat Tujuan penelitian, melakukan pengumpulan data, melakukan Data preprocessing, membuat Data Training dan Data testing. Pengolahan Data untuk membuat Model Klasifikasi penjualan rumah.

3.4 Populasi dan Sampel

Data yang akan dipakai untuk penelitian ini diambil atau *crawling* dari situs *e-commerce* rumah123.com dengan berbentuk *Excel Worksheet*. Data terdiri dari penjualan rumah di kota Bandung.

3.5 Teknik Analisis Data

Dalam menganalisis data, penulis menggunakan metode klasifikasi dan deskripsi.^[10]

3.5.1 Data Preprocessing

Pembersihan data atau data cleaning dapat diaplikasikan untuk menghilangkan noise dan memperbaiki data inkonsisten.

3.5.2 Teknik Klasifikasi

Klasifikasi adalah suatu bentuk analisis data yang mengestrak model untuk menggambarkan kelas data yang penting, seperti model yang disebut klasifikasi, tujuannya untuk memprediksi kategori (diskrit, *unordered*) label kelas.^[13]

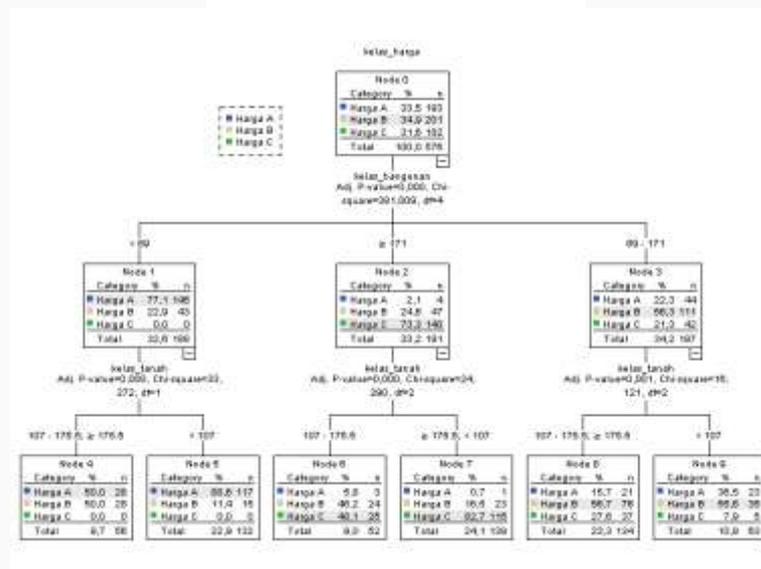
3.5.3 Confusion Matrix

Metode klasifikasi akan dilakukan evaluasi pertama pada bagian akurasi dari hasil klasifikasi. Akurasi sebuah klasifikasi berpengaruh terhadap performa dari suatu klasifikasi.^[6]

4. Hasil Penelitian

4.1 Decision Tree

Berikut adalah hasil dari model Decision Tree, Setelah data melewati tahap preprocessing dengan mengubah data kontinu menjadi diskrit dan menghilangkan data yang tidak diperlukan maka data sudah dapat diubah menjadi model Decision Tree. Variabel independen atau variabel yang dijadikan target yaitu kelas_harga. Variabel yang digunakan lainnya yaitu kelas_bangunan, kelas_tanah, kamar mandi, kamar tidur. Data Training yang digunakan sebanyak 70% data dan Data Testing sebesar 30%. variabel yang paling berpengaruh dalam menentukan kelas_harga adalah kelas_bangunan.

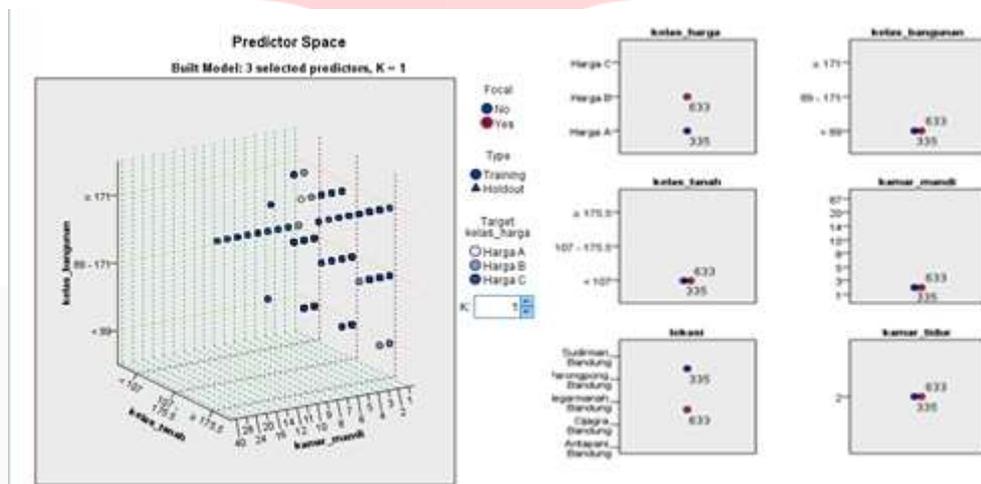


Gambar 2 Model Decision Tree

Sumber: Olahan Penulis

4.2 K-NN

Sesuai algoritma nearest neighbor(k=1) untuk memilah data agar terlihat kemiripan dari datanya. Dari gambar dibawah ini bisa dilihat Kelas A mempunyai warna Biru keputihan, Kelas B Biru dan Kelas C mempunyai warna Biru Tua. Bentuk data training dengan pola Bulat sedangkan Testing dengan pola segitiga. Dari variable-variabel yang digunakan yang paling berpengaruh adalah kelas_bangunan untuk menentukan kelas_harga, diikuti dengankelas_tanah dan kamar_mandi.



Gambar 3 Model KNN

Sumber: Olahan Penulis

Peneliti mengambil contoh dengan melihat pada data 633. Data ke 633 mempunyai kemiripan data yaitu data ke 335. Dari gambar ini terlihat, data 633 & 335 memiliki kesamaan pada kelas_tanah, kelas_bangunan dan kamar_tidur sedangkan memiliki perbedaan mencolok pada kelas_harga dan lokasi. Pada data 633 termasuk kategori Harga B sedangkan 335 masuk kategori harga A begitu juga dilihat pada lokasi.

4.3 Confusion Matrix

Confusion Matrix (Decision Tree) merupakan evaluasi pada model Decision Tree. Dari perhitungan evaluasi persentase akurasi pada model Decision Tree berhasil memberikan akurasi sebesar 75,10% sedangkan KNN memiliki presentase akurasi sebesar 71,30%.

Tabel 1 Evaluasi Akurasi Confusion Matrix

Decision Tree	75,10%
KNN	71,30%

Sumber: Olahan Penulis

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan hasil yang telah didapatkan dalam bab sebelumnya, peneliti menyimpulkan:

1. Rumah yang paling banyak terjual ada didaerah Arcamanik, Antapani, Soekarno-Hatta dan Buah Batu. Rata-rata harga pada rumah paling banyak terjual yaitu Arcamanik senilai 800 juta, Antapani senilai 1,1 Miliar, Soekarno-Hatta 800 juta dan Buah Batu senilai 1,2 Miliar. Rata-rata luas tanah pada rumah paling banyak terjual yaitu Arcamanik 143 m², Antapani 123 m², Soekarno-Hatta 123m² dan Buah Batu 163 m². Rata-rata luas bangunan pada rumah paling banyak terjual yaitu Arcamanik 106 m², Antapani 138m², Soekarno-Hatta 126 m² dan Buah Batu 156 m².
2. Analisis model prediksi terhadap data penjualan rumah di bandung dengan metode Decision Tree memiliki akurasi sebesar 75,10% dan metode KNN sebesar 71,30%.

5.2 Saran

5.2.1 Aspek Akademis Untuk Penelitian Selanjutnya

Penelitian ini menggunakan pendekatan metode klasifikasi Decision Tree dan K-Nearest Neighbor. Selanjutnya dapat dilakukan bagaimana implementasi pada bisnis lainnya dan/atau melakukan penelitian dengan data yang *unsupervised* seperti *Clustering*.

5.2.2 Aspek Praktis

Penelitian ini dapat dijadikan bahan referensi visualisasi data kepada masyarakat umum untuk membeli ataupun menjual rumah di kota Bandung dan untuk para pebisnis dapat dijadikan pengetahuan untuk melakukan bisnis properti pada daerah yang paling populer.

Daftar Pustaka

- [1] *Bandung Kota Terbaik untuk Bisnis Properti*. (2016). Diambil kembali dari Property Bandung: <http://www.property-bandung.com/kpr/8-artikel/2-bandung-kota-terbaik-untuk-bisnis-properti.html>

- [2] Bhaskara, I. L. (2016, Desember 12). *Geliat e-Commerce Berkah Industri Logistik*. Diambil kembali dari tirto.id: <https://tirto.id/geliat-e-commerce-berkah-industri-logistik-b854>
- [3] *Ini 'Surga' Baru Bisnis Properti di Bandung*. (2016, Desember 11). Diambil kembali dari Okezone Finance: <http://economy.okezone.com/read/2016/12/09/470/1563126/ini-surga-baru-bisnis-properti-di-bandung>
- [4] *Internet Users in Asia June 2016*. (2016). Diambil kembali dari Internet World Stats: <http://www.internetworldstats.com/stats3.htm>
- [5] Zaki, M. J., & Meira Jr, W. (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*. Cambridge University Press.
- [6] Han, J. and Kamber, M, 2006, "Data Mining Concepts and Techniques Second Edition". Morgan Kauffman, San Francisco.
- [7] Bhardwaj, B. K., & Pal, S. (2012). Data Mining: A prediction for performance improvement using classification. arXiv preprint arXiv:1201.3418.
- [8] Lei, Y. (2013). Web Data Mining Technology and Instrument Research. In Proceedings of the 2nd International Conference on Green Communications and Networks 2012 (GCN 2012): Volume 2 (pp. 231-237). Springer Berlin Heidelberg.
- [10] Larose, Daniel T., dan Larose, Chantal D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining Second Edition*. New Jersey: John Wiley & Sons Inc.
- [11] Dumbill, E. (2012). Planning for big data. " O'Reilly Media, Inc.". Didapat dari : http://cdn.oreillystatic.com/oreilly/radarreport/0636920025559/Planning_for_Big_Data.pdf
- [12] Laudon, K. C., & Traver, C. G. (2012). *E-commerce 2012: Business, Technology, Society*.
- [13] Han, Jiawei., Kamber, Micheline., dan Pei, Jian. (2012). *Data Mining Concepts and Techniques Third Edition*. Waltham: Elsevier.
- [14] Maimon, O., & Rokach, L. (2008). *Data mining with decision trees: theory and applications*