

K-Nearest Neighbor Implementation in the application of Appraisal Comparison Data Search

Doni Winarso
Department of Information Systems
Universitas Muhammadiyah Riau
Pekanbaru, Indonesia
doniwinarso@umri.ac.id

Edo Arribe
Department of Information Systems
Universitas Muhammadiyah Riau
Pekanbaru, Indonesia
edoarribe@umri.ac.id

Abstract - Appraisal is an activity to estimate the price of an object. This activity is often used to get a fair price for a property. The approach that is often used in property valuation is to compare objects that are valued with market data as a comparison. This study aims to obtain comparative data on appraisal. The method used is the K-Nearest Neighbor (K-NN) and euclidean distance classification method to find the proximity of the object being assessed with market data around the object to be assessed. The finding of this study is in the form of a comparative data search application using the K-NN method with a total $K = 3$.

Keywords - Appraisal Comparison Data; K-Nearest Neighbor; simple property

I. INTRODUCTION

Appraisal is the estimated price of a property. Appraisal activities are often carried out by the Office of Public Appraisal Services (KJPP). Appraisal is also used by internal banks to find out how much the value of the property to be used as collateral for financing. According to the regulation of the Minister of Finance of the Republic of Indonesia, there are 3 types of property valuations carried out by appraisers, including simple property valuations, property valuations and business valuations [1]. In valuing simple property, there are 4 approaches that can be used by appraisers, namely the market approach, the cost approach, the income approach and the price hedonic approach [1], [2]. From these 4 approaches, the market data approach is often chosen by the appraiser to get a fair price for a property.

Appraisal using the market data approach is done by considering the sale of similar or substitute property and related market data, and producing estimated values through a comparison process. In general, the property being valued (the object of valuation) is compared with comparable property transactions, both those that have occurred and properties that are still in the bidding stage of the sale and purchase process [4]. The approach to the market data comparison method assumes that properties that have a high degree of similarity will have similarities in price and benefits.

This study will discuss the search for simple property comparison data in the form of land and buildings that refer to the Regulation of the Minister of Finance (PMK) number 101 of 2014 [1]. Search for simple property comparison data so far is still done using the experience and intuition of the assessor. Subjectivity arising from the evaluator has a big influence on the determination of comparison data, so that the results of the assessment are not consistent and raise questions for users of the report [5].

The concept of comparing data search is through comparative analysis of the nearest neighbors, and the degree of similarity of 2 objects. This is in accordance with the basic concept of the K-Nearest Neighbor (K-NN) method where this method classifies data based on the degree of similarity [6]. The K-NN method is one of the 10 most popular data mining methods [7]. This method is simple, reduces training time and has good performance [8].

Many fields and problems have been solved using the K-NN method, including those for prediction [9], [10] classification and identification [11], [12] and many other related studies using these methods.

In this study, K-NN is used to classify comparable data that is suitable or similar to the object property to be assessed. For each object being assessed, at least 3 comparative data are needed from surrounding locations that have similar criteria.

II. RESEARCH METHODOLOGY

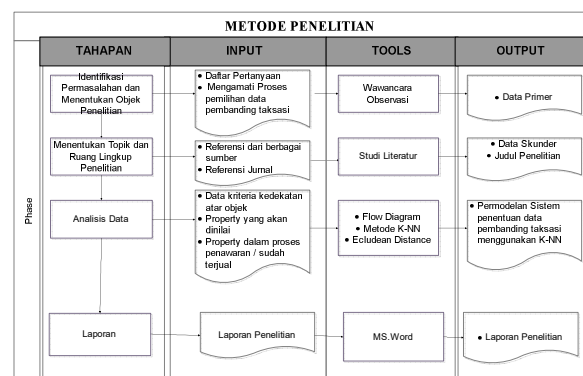


Figure 1. Reasearch Methodology

This research begins with the identification of the problems that exist in the process of valuing a simple property, especially at the stage of selecting comparative data to be used to assess a simple property. Identifying the problem and determining the object of research was done by interviewing and observing, namely observing a simple property valuation process. At this stage, the research data obtained in the form of criteria in the determination of comparative data valuation of simple property.

The next process is to look for references in the form of journals, books, reports related to the process of selecting appraisal comparison data and a simple property valuation process. This reference is used to support this research.

The data used in this study are 45 property data that are in the process of bidding and some data that have been sold in the city of Pekanbaru. Initial data was then carried out the data cleaning process. Data cleaning was done by looking for criteria that affect the level of similarity between the comparison data with the data to be assessed. The criteria were then valued between 0 and 1 using the following formula.

$$W_j = \frac{W_0}{\sum W_0} \dots \dots \dots (1)$$

W_j = Value improvement

W_0 = Initial value

$\sum W_0$ = Total value

The analysis was performed using the K-NN method to find similarities with the nearest neighbors or the most similar to the property data to be assessed. In calculating the distance for the degree of similarity, the euclidean distance formula is used, with the following formula [13]:

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \dots \dots \dots (2)$$

Notes:

d_{ij} = Degree of difference

n = Number of vectors

x_{ik} = Vector citra input

x_{jk} = Comparison vector citra /output.

III. RESULTS AND DISCUSSION

At this stage an analysis of data on property sales in the city of Pekanbaru was carried out to get the criteria as a reference to calculate the level of similarity between objects to be assessed with the nearest comparable data. The architecture of this research is as follows:

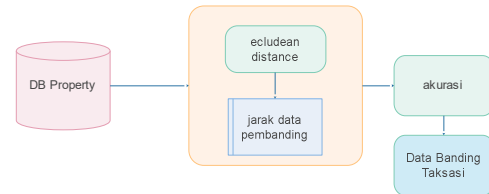


Figure 2. System Architecture

Data Collection

This research uses primary data and secondary data. Perimer data in this study were taken directly in the field in the form of house properties that are being offered and house properties that have been sold. Furthermore, secondary data in this study are data on home property sales in Pekanbaru which are offered via the olx.co.id web marketplace and rumah123.com. The data used are data on the sale of home properties offered or sold in the January 2018 to July 2019 period.

Data Analysis

The first analysis is to look for criteria to be used in calculating closeness between objects. From the results of interviews and observations conducted, criteria and sub-criteria and their value were obtained. Value calculation uses formula number 1. Criteria data for sub-criteria and value are presented in the following table

Table.1 Appraisal Data Criteria

No	Criteria	Sub Criteria	Value
1.	Type of Asset	Vacant Land	0.17
		Housing	0.33
		Shop house	0.50
2.	Legality	Cutting down	0.10
		SKGR	0.20
		SHGB	0.30
		SHM	0.40
3.	Land Size	< 100	0.04
		100 s/d 199	0.07
		200 s/d 499	0.11
		500 s/d 999	0.14
		1000 s/d 4999	0.18
		5000 s/d 6999	0.21
		>7000	0.25
4.	Distance to Security Center	<100	0.07
		100 s/d 499	0.13
		500 s/d 999	0.20
		1000 s/d 2999	0.27
		>3000	0.33
5.	Front Path Width of the Object	<1	0.07
		1 s/d 2	0.13
		3 s/d 4	0.20
		5 s/d 6	0.27
		>6	0.33
6.	Road Quality	Asphalt	0.10
		Hardening	0.20
		Cement	0.30
		Soil	0.40

		Others	0.30
		None	0.40
7.	Ease of Transportation	Very Easy	0.07
		Easy	0.13
		Quite Easy	0.20
		Difficult	0.27
		None	0.33
8.	Surrounding Area	Very Prestigious	0.17
		Quite Prestigious	0.33
		Less Prestigious	0.50
9.	Land Development Trend	Very Fast	0.07
		Faster	0.13
		Do not develop to	0.20
		Tend decrease	0.27
		Decrease dramatically	0.33
10.	Foundation	Reinforced concrete cast	0.17
		River Stone	0.33
		Natural Brick	0.50
11.	Floor	Granite	0.17
		Ceramics	0.33
		Aci Cement	0.50
12.	Number of Bathrooms	1	0.05
		2	0.10
		3	0.14
		4	0.19
		5	0.24
		>5	0.29
13.	Number of Bedrooms	1	0.05
		2	0.10
		3	0.14
		4	0.19
		5	0.24
		>5	0.29
14.	Ceiling	Sunda Plafon	0.07
		Gypsum	0.13
		Plywood	0.20
		Others	0.27
		None	0.33
15.	Roof	Tile	0.10
		Multiroof	0.20
		Galvalum roof	0.30
		Asbes	0.40
16.	Water Sources	Artesian Well	0.10
		Ring Well	0.20
		PAM	0.30
		None	0.40
17.	Electricity Sources	PLN	0.10
		Genset	0.20

18.	Garden	Any	0.33
		None	0.67
19.	Canopy	Any	0.33
		None	0.67
20.	Fence	Any	0.33
		None	0.67
21.	Carport	Any	0.33
		None	0.67
22.	Building Size	< 36	0.07
		36 s/d 38	0.13
		39 s/d 45	0.20
		46 s/d 60	0.27
		60 s/d 70	0.33
		>70	

Testing is done by calculating the proximity of the training data with a data property to be sold. The distance calculation uses the euclidean distance formula in formula number 2. The similarity results are obtained by displaying K = 3 of the proximity of the distance for each data.

Table 2. Proximity to test data

No	Data Property	Proximity
1	x1	0,61
2	x2	0,67
3	x3	0,63
4	x4	0,58
5	x5	0,45
6	x6	0,61
7	x7	0,58
8	x8	0,43
9	x9	0,61
10	x10	0,61
11	x11	0,58
12	x12	0,55
13	x13	0,46
14	x14	0,51
15	x15	0,46
16	x16	0,53
17	x17	0,54
18	x18	0,53
19	x19	0,52
20	x20	0,53
21	x21	0,61
22	x22	0,46

23	x23	0,17
24	x24	0,44
25	x25	0,61
26	x26	0,10
27	x27	0,52
28	x28	0,49
29	x29	0,50
30	x30	0,57
31	x31	0,56
32	x32	0,56
33	x33	0,62
34	x34	0,52
35	x35	0,51
36	x36	0,56
37	x37	0,50
38	x38	0,60
39	x39	0,47
40	x40	0,54
41	x41	0,46
42	x42	0,47
43	x43	0,48
44	x44	0,51

From the calculations in table 1, it is found that the proximity of the 3 closest neighbors is data number 26, 23 and data number 8. The distance for each nearest neighbor is presented in the following table.

Table 3. The closest neighbor is K=3

Data Number	Data Property	Proximity
26	x26	0,10
23	x23	0,17
8	x8	0,43

Implementation

The system was developed using PHP programming language and design using bootstrap 3.3.7. The local server uses the XAMPP application which includes Apache server and MySQL Data Base Management Systems (DBMS). The process of finding appraisal comparison data was carried out through the following user interface (UI).

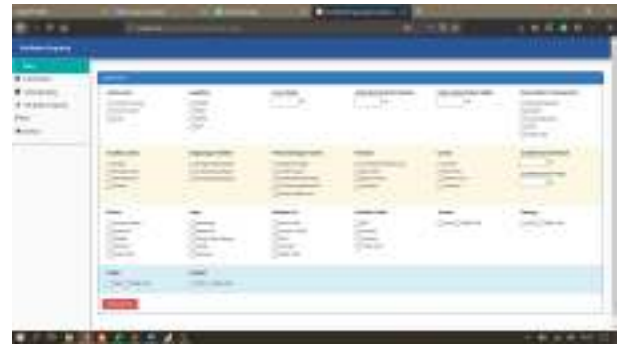


Figure 3. UI Process of Appraisal Comparison Data



Figure 4. UI Calculation Results of Appraisal Comparison Data

IV. CONCLUSION

From the research conducted by KNN, it can be used to find and determine appraisal comparative data. But KNN requires a long processing time because KNN will calculate the distance of each data with each data having 22 criteria.

ACKNOWLEDGEMENT

The series of thanks to those who have played a role in this research, especially to the Ministry of Research, Technology and Higher Education that have funded this research through contract number: 7/E/KPT/2019 19 February 2019. Our gratitude also goes to the Institute for Research and Community Services of Universitas Muhammadiyah Riau.

REFERENCES

- [1] M. K. R. Indonesia, "PMK 101 Tahun 2014 Tentang Penilai Publik." Jakarta, 2014.
- [2] S. Peterson and A. B. Flanagan, "Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal," *J. Real Estate Res.*, vol. 31, no. 2, pp. 147–164, 2009.
- [3] W. J. McCluskey, M. McCord, P. T. Davis, M. Haran, and D. McIlhatton, "Prediction accuracy in mass appraisal: A comparison of modern approaches," *J. Prop. Res.*, vol. 30, no. 4, pp. 239–265, 2013.

- [4] K. Penyusun, S. Penilaian, M. Profesi, and P. Indonesia, "Petunjuk Teknis SPI 306," 2015.
- [5] I. C. Yeh and T. K. Hsu, "Building real estate valuation models with comparative approach through case-based reasoning," *Appl. Soft Comput. J.*, vol. 65, pp. 260–271, 2018.
- [6] D. Winarso and E. Arribe, "Seleksi Pegawai dan Dosen UMRI Berbasis E-Recruitment," *J. Teknol. Inf. Komun. Digit. Zo.*, vol. 8, no. 2, pp. 71–80, 2017.
- [7] X. Wu and V. Kumar, *The Top Ten Algorithms in Data Mining*. New York: CRC Press, 1382.
- [8] and O. O. O. J. F. Ajao, D. O. Olawuyi, "Yoruba Handwritten Character Recognition using Freeman Chain Code and K-Nearest Neighbor Classifier," *J. Teknol. dan Sist. Komput.*, vol. 6, no. 4, p. 129, 2018.
- [9] G. O. F. Mustakim, "Algoritma K-Nearest Neighbor Classification," vol. 13, no. 2, pp. 195–202, 2016.
- [10] M. Sharma, "Generalized K-Nearest Neighbour Algorithm- A Predicting Tool," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 11, pp. 1–4, 2013.
- [11] J. E. Vidueira Ferreira, C. H. S. da Costa, R. M. de Miranda, and A. F. de Figueiredo, "The use of the k nearest neighbor method to classify the representative elements," *Educ. Química*, vol. 26, no. 3, pp. 195–201, 2015.
- [12] F. S. Ni'mah, T. Sutojo, and D. R. I. M. Setiadi, "Identifikasi Tumbuhan Obat Herbal Berdasarkan Citra Daun Menggunakan Algoritma Gray Level Co-occurrence Matrix dan K-Nearest Neighbor," *J. Teknol. dan Sist. Komput.*, vol. 6, no. 2, p. 51, 2018.
- [13] S. Wurdianarto, S. R. Wurdianarto, S. Novianto, and U. Rosyidah, "Perbandingan Euclidean Distance Dengan Canberra Distance Pada Face Recognition," *Techno.Com*, vol. 13, no. 1, pp. 31–37, 2014.