

Analisa Pengaruh Objek Spasial Terhadap Harga Sembako di Kota Bandung

Analysis of The Influence of spatial objects on the Price of Basic Food in Bandung

Seno Anunggroho

Prodi S1 Teknik Informatika, Fakultas Informatika, Universitas Telkom

Ssenoo19@gmail.com

ABSTRAK

This research discusses the analysis of influence of spatial object that exists around the market in Bandung against the price of basic food. To determine the required price data analysis is complete food prices, Basic food data obtained from the Office of Industry and Trade of West Java and will only take sampling from 5 markets in the city of Bandung that is Kiara Condong Market, Baru Market, Sederhana Market , Andir Market Dan Kosambi Market. The theory to be used in this research is Iterative Dichotomiser 3 one of theory from decision tree,the author use this method because the decision tree can find hidden relationships or patterns between variables .With this analysis we can know how the influence of the object on the price of basic food in the city of Bandung.Kata Keyword : Decision Tree, ID3,Basic Food

1. Introduction

Basic food is a major requirement for all people whether it is for everyday needs or even for businesses related to basic food. Basic food prices could fluctuate and difficult to predict how large the increase or decline in basic food prices

Many factors affecting the price of basic foods include climate factors, foreign factors for imported basic foods, fuel increase, distribution factor and speculative factors. For the cost price of this spatial factor, it is not yet known how the effect on the price of basic foods. Research on the influence of the price of basic foods from spatial factors is still rare and even no one has done. The meaning of the spatial factor here is the facility that is located around the market. Therefore, to know the facilities that affect the price of basic food of analysis needed to determine how the influence of spatial objects on the price of basic food. Here I take a case study in Bandung city by taking sampling in five markets in Bandung city according to data from the Office of Industry and Trade of West Java with market details of Kiara Condong Market, Pasar Baru Market, Sederhana Market, Andir Market and Kosambi Market.

Analysis of the influence of spatial objects that affect the basic foods needs a method where the prediction methods that can be used should generate predictions with a high degree of accuracy. The author will use Decision Tree by using ID3 (Iterative Dichotomiser 3) method where this method can find the hidden relation between variables.

2. Literatur

2.1. Decision Tree

The most commonly used method of data mining is the decision tree. The decision tree method transforms a very large fact into a decision tree that represents the rule. The concept of a decision tree is to transform data into decision trees and decision rules. Data in a decision tree is usually represented in a table with attributes and records. The attribute states a parameter created as a criterion in the building of a tree. The process on the decision tree is to convert the data form into a tree model, change the tree model into a rule, and simplify the rule. The main benefit of using decision trees is its ability to break down complex decision making processes to be simpler so that decision makers will better interpret the solution of the problem.[2]

2.2. ID3

ID3 begins with the original set S as the root node. In each algorithm iteration, iterating through each unused attribute of the set S and computing the entropy (S) Information Gain of all custom attributes. Then, select the attribute that has the highest Information Gain[3]. Step work of ID3 algorithm is as follows;

• Calculate entropy and information gain of each attribute with the rmus as follows The purpose of the calculation of information gain is to choose which attribute will be used as a branch on the formation of a decision tree by selecting the attribute that has the largest information gain value

$$Entropy(S) = -P + \log_2 P + -P - \log_2 P - \quad [4]$$

S = space (data) sample used for training.

$P +$ = positive-scaled numbers (support) in the sample data for certain criteria.

$P -$ = negative-collected amount (not supported) in the sample data for certain criteria.

$$Gain(S, A) = Entropy(S) - \sum \frac{|S_v|}{|S|} Entropy(S_v) \quad [4]$$

S = space (data) sample used for training.

A = attribute.

V = a possible value for attribute A .

Value (A) = possible set for attribute A .

S_v = Number of samples for value V .

S = Total number of sample data.

Entropy (S_v) = entropy for samples that have a value of V .

- A node shape that contains the attributes of the calculated informatin gain.
- Repeat the process of calculating the information gain until all data has been included in the same class, the attribute that has been used will not be calculated again information gain. Id3 will stop if all attributes have classified training set.[3]

2.3. Preprocessing

Data preprocessing is one step of eliminating problems that can occur and interfere with the results rather than process[5]. Step Preprocessing is,

- Cleaning Data: Data is cleared through processes such as filling in missing values, facilitating noisy data, or resolving data inconsistencies.
- Data Integration: Data with different representations are combined and conflicts in data are solve
- Data Transformation: Data is normalized, combined and general.
- Data Reduction: This step aims to present a diminished data representation in the data warehouse.
- Data Discretization: Involves the reduction of a number of continuous attribute values by dividing the range of the attribute interval.[6]

2.4. Data Mining

Data mining involves theory, methodology, and computer systems for the extraction of knowledge or mining of large amounts of data[7]. Data mining is necessary to do especially in managing the data is very large to facilitate the activities of recording a transaction and for data warehousing process in order to provide accurate information for users, in data mining there are also classification and clustering where classification is the processing yag aimed to find a model or function and rules that explain and

characterize the concept or class of data, for a particular interest, while Ccustering is used for grouping data based on similarity in the object data and minimize the similarity to other clusters.[8] Data mining is also known by other names such as Knowledge discovery (mining) in databases (KDD), knowledge extraction, data analysis and business intelligence and is an important tool for manipulating data for the presentation of information as needed users with a view to assist in the analysis of behavioral observation collections.[9]

3. Methodology and Result

3.1. System Decsription

The system to be created is a system that can analyze the influential spatial objects and how the influence of these objects on the price of basic food in the city of Bandung. The analysis will be done by determining the pattern of all the possibilities formed from the price of basic foods and objects around it, so it is expected to produce the best rule. Analysis of the influence of spatial objects on the basic food will use one of the decision tree method of ID3.

3.2. Clasification With Decision Tree

At this stage will be determined what objects are influential and how the influence of the object on the price of groceries in the city bandung.Object to be analyzed are schools, restaurants, hospitals, health centers, tours, offices, housing and hotels.Determine the analysis of the increase in the price of sembako from the spatial side using one of the decision tree's algorithm Id3. I use decison tree because decision tree can find fact or relationship hidden between class of input variable in this case facility, distance and type of sembako with target class that is price sembako.Id3 have two main concept that is entropy and information gain, information gain function to determine Which attribute will be the branch of the decision tree by selecting the highest information gain, while entropy is involved in the calculation of the information gain, following the example of entropy and information gain.

Item	Facility	Distance	Price
Gula	School	S<1	High
Gula	School	2≤S<3	High
Kentang	Hotel	S<1	High
Kentang	Hotel	S<1	Low
Gula	Hotel	1≤S<2	Low
Entang	Hotel	1≤S<2	High
Gula	Hotel	S<1	High
Kentang	School	1≤S<2	High
Kentang	Hotel	2≤S<3	Low
Kentang	School	2≤S<3	High

Based on the table when viewed from the price column found that there are 6 high prices and 4 low prices then entropy is

$$Entropy(S) = -P + \log_2 P + - P - \log_2 P -$$

$$Entropy(6+, 4-) = - \left(\frac{7}{10}\right) \log_2 \left(\frac{7}{10}\right) - \left(\frac{3}{10}\right) \log_2 \left(\frac{3}{10}\right) = 0.88$$

Whereas, for the information gain of the item column is

Values Item

S Kentang=[4+,2-]

S Gula=[3+,1-]

$$Gain(S, A) = Entropy(S) - \sum \frac{|Sv|}{|S|} Entropy(Sv)$$

$$Gain(S, Item) = 0.88 - \left(\frac{6}{10}\right) 0.91 - \left(\frac{4}{10}\right) 0.81 = 0.11$$

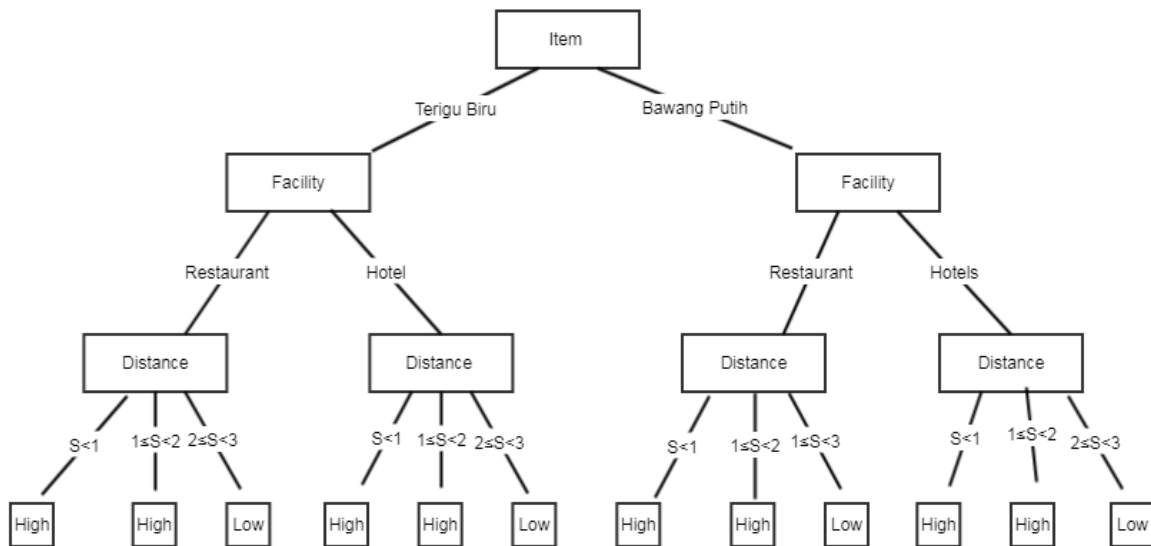
3.3 Result

After analyzing the patterns obtained ID3 method influence spatial relationship between objects, types of groceries with food prices, The spatial object here involves the type of facility, the distance of the facility to the market and the price of the basic needs, The four variables have a relationship in determining the pattern obtained, here are some results of the pattern obtained

Item	Facilty	Distance(KM)	Price
Terigu Cakra	Restaurant	2 <= Distance <= 3	High
Terigu Cakra	Restaurant	1 <= Distance <2	High
Terigu Cakra	Restaurant	Distance < 1	High
Cabai Merah Biasa	School	2 <= Distance <= 3	Low
Cabai Merah Biasa	School	1 <= Distance <2	Low
Cabai Merah Biasa	School	Distance < 1	High
Kentang	Office	2 <= Distance <= 3	Low
Kentang	Office	1 <= Distance <2	High
Kentang	Office	Distance < 1	High
Ikan Asin	School	2 <= Distance <= 3	Low
Ikan Asin	School	1 <= Distance <2	Low
Ikan Asin	School	Distance < 1	Low
Telur Broiler	Hotels	2 <= Distance <= 3	Low
Telur Broiler	Hotels	1 <= Distance <2	Low
Telur Broiler	Hotels	Distance < 1	High
Telur Broiler	Restaurant	2 <= Distance <= 3	Low
Telur Broiler	Restaurant	1 <= Distance <2	Low
Telur Broiler	Restaurant	Distance < 1	Low
Susu Bendera	Tours	2 <= Distance <= 3	Low
Susu Bendera	Tours	1 <= Distance <2	Low
Susu Bendera	Tours	Distance < 1	High
Jagung Pipilan	Office	2 <= Distance <= 3	Low
Jagung Pipilan	Office	1 <= Distance <2	High
Jagung Pipilan	Office	Distance < 1	High
Terigu Biru	Hotels	2 <= Distance <= 3	High
Terigu Biru	Hotels	1 <= Distance <2	High
Terigu Biru	Hotels	Distance < 1	Low
Terigu Biru	Restaurant	2 <= Distance <= 3	High
Terigu Biru	Restaurant	1 <= Distance <2	High
Terigu Biru	Restaurant	Distance < 1	Low
Bawang Putih	Hotels	2 <= Distance <= 3	Low
Bawang Putih	Hotels	1 <= Distance <2	High

Bawang Putih	Hotels	Distance < 1	High
Bawang Putih	Restaurant	2 <= Distance <= 3	Low
Bawang Putih	Restaurant	1 <= Distance <2	High
Bawang Putih	Restaurant	Distance < 1	High

For illustration of the decision tree, the following will be exemplified illustration of the tree for terigu biru dan bawang putih



4. Conclusion

To get the pattern of influence of spatial object in this research the method used is decision tree by using Id3 algorithm. Facilities taken amounted to 8 facilities namely Hospital, Puskesmas, Hotel, Office, School, Housing, Restaurant and tourist attractions. While the market taken as sampling there are five that is Pasar Baru, Kiara Condong Market, Andir Market, Kosambi Market and Sederhana Market. From the facility and the market will be searched the distance from each facility to the market to then be sought pattern that can know the influence of the object to the price of basic food, After analyzing the four variables turn out the four variables are mutually related, and ID3 method can find the rules of the relation.

Bibliography

[1] *Konsep Pohon Keputusan ID3 dan C4.5*. <http://pohonkeputusan.com/konsep-pohon-keputusan-id3-dan-c4-5>, 2015

[2] *Sayad, S Decision Tree - Classification*, 2010

[3] *Baldwin, P., David, P ID3 (Iterative Dichotomiser 3)*, <http://debategraph.org/Details.aspx?nid=305543/>

[4] *Selman, B. Machine Learning: Decision Trees*, R&N 18.3

[5] Han, J., and Kamber, M *Morgan Kaufman Publisher. Data Mining Concept and Technique., 2006, San Fransisco*.

[6] *Techopedia inc. Data Preprocessing*. <https://www.techopedia.com/definition/14650/datapreprocessing,2017>

[7] *Telir, S. Kanika, P. International Journal of Advanced Research in Computer Science and Software Engineering. A Survey on Decision Tree Based*

- Approaches in Data Mining, 2015*
- [8] Devianti,S. Jajui,M,. *International Journal of Advanced Research in Konferensi Nasional Informatika (KNIF) 2015 Integrasi Metode Klasifikasi Dan Clusteringdalam Data Mining, 2015*
- [9] gsbipb , *Pengertian dan Konsep Data Mining <http://gsbipb.com>,2013*