SPATIAL DATA MINING FOR POOR VILLAGE CHARACTERIZATI AT WEST JAVA AND BANTEN- INDONESIA

H.A. Adrianto [1],

[1] Department of Computer Science, Faculty of Mathematics and Natural Sciences, Institut Pertanian Bogor
agung@ipb.ac.id

ABSTRACT:

Data mining, also popularly referred to as Knowledge discovery in databases (KDD), is the automated or convenient extraction of pattern representing knowledge implicitly stored in large databases, data warehouses, and other massive information repositories. Spatial data mining deal with knowledge discovery in spatial databases. One of spatial data mining task is characterization – to find a description of the spatial and non-spatial properties which are typical for the target objects but not for the whole database. In this research we developed a system prototype which can be implement spatial data mining technique to discover poor village characteristics at West Java and Banten. The prototype consist of four components: pre-processing module, spatial data mining module, database module and visualization module. Spatial data mining module consist of neighbourhood graph, predicate filter, neighbourhood index and spatial characterization algorithm. From this research we discovered several factor related with village poverty i.e. road type, the existence of newspaper subscribers, distance form health facility and type of waste disposal facility.

KEYWORDS: Data Mining, Spatial Characterization, Poor Villages, West Java, Banten

1. INTRODUCTION

Spatial data mining deal with knowledge discovery in spatial databases. One of spatial data mining tasks is characterization – finding a description of the spatial and non-spatial properties which are typical for the target objects but not for the whole database. Ester *et. al* (1998) discovered spatial description on Bavarian census database, informally founded rule states that "retires prefer somewhat rural areas close to the mountains".

Our research use spatial characterization to discover poor village characteristics in West Java and Banten, Indonesia. The aim of this research is to develop a prototype which can do characterization technique for West Java and Banten Potential of Village Statistics year 2003. Banten itself is a new province apart from West Java at 2000.

2. METHODOLOGY

2.1 Spatial Characterization

Ester *et.al* (2001) state that the task of characterization is to find a compact description for a selected subset of the database. They define a spatial characterization of a set of target objects with respect to the database containing them as a description of the spatial and non-spatial properties which are typical for the target objects but not for the whole database. Spatial characterization use the relative frequencies of the non-spatial attribute values and the relative frequencies of the different object types as the interesting properties.

Let $G_{neighbour}^{DB}$ be a neighbourhood graph and *targets* be a subset of *DB*. Let *freq^s(prop)* denote the number of occurrences of the property *prop* in the set *s* and let *card(s)* denote the cardinality of *s*. The *frequency factor* of *prop* with respect to *targets* and *DB* is denoted by

$$f_{t\arg ets}^{DB}(prop) = \frac{freq^{t\arg ets}(prop)}{card(t\arg ets)} / \frac{freq^{DB}(prop)}{card(DB)}$$

Let *significance* and *proportion* be real numbers and let *max-neighbours* be a natural number. Let *Neighbours* $^{i}_{G}(s)$ denote the set of all objects reachable from one of the elements of *s* by traversing at most *i* of the edges of the neighbourhood graph *G*.

Then, the task of *spatial characterization* is to discover each property *prop* and each natural number $n \leq$ *max-neighbours* such that

(1) the set *objects* = $neighbours_G^n(t\arg ets)$ as well as

(2) the set *objects* = $neighbours_G^n(\{t\})$

for at least *proportion* many $t \in target$ satisfy the condition:

$$f_{objects}^{DB}(prop) \begin{array}{l} \geq significance \\ \text{or} \\ \leq \dfrac{1}{significance} \end{array}$$

The generated rule has the following format:

$$target \Rightarrow p_1\,(n_1, freq\text{-}fac_1) \wedge \ldots \wedge p_k\,(n_k, freq\text{-}fac_k).$$

This rule means that for the set of all targets extended by n_i neighbours, the property p_i is *freq-fac$_i$* times more (or less) frequent than in the database. So that the general spatial characterization algorithm works as follows (Ester *et.al* 2001) :

```
discover-spatial-characterization (graph $G_r^{DB}$; set of objects target; real significance, proportion; integer max-neighbours)

initialize the set of characterizations as empty;
initialize the set of regions to targets;
initialize n to 0;
calculate $frequency^{DB}(prop)$ for all prop = (attribute, value) in DB;

while $n \le$ max-neighbours do
    for each attribute of DB and for the special
        attribute object type do
        for each value of attribute do
            calculate $frequency^{region}(prop)$ for
            property prop = (attribute, value);
            if $f_{region}^{DB}(prop) \ge significance$ or
               $f_{region}^{DB}(prop) \le 1/significance$
            then
                add (prop, n, $f_{region}^{DB}(prop)$) to the set
                characterizations;
    if n < max-neighbours then
        for each object in region do
            add neighbours ($G_r^{DB}$, object, TRUE) to
            region;
    increment n by 1;
extract all tuples (prop, n, $f_{region}^{DB}(prop)$) from
charactrizations which are significant in at least
proportion of the regions with n extensions;

return the rule generated from these
characterizations;
```

2.2 Prototype Architecture

Spatial characterization prototype consists of four components, i.e. pre-processing, spatial data mining, database and visualization. Dependability among those components shown at Figure 1.

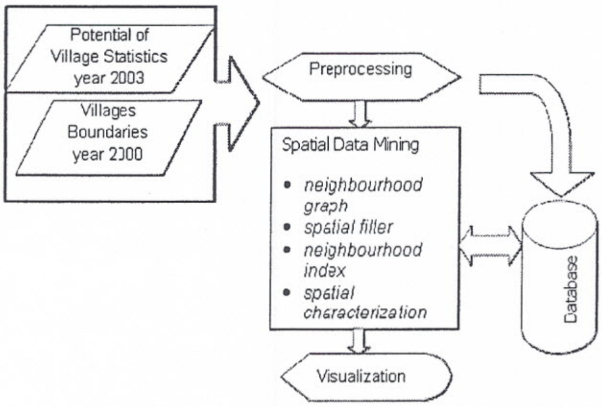


Figure 1. Prototype Architecture

Data used in this research are digital map of West Java Province year 2000 and Potential of Village Statistics (PODES) year 2003. The map consist of 7327 villages boundary whereas PODES data records attributes for 7234 villages. After data cleaning process we have 6248 villages match (86%).

Generally, PODES data contain 744 attributes i.e characteristics, population, housing and environment, education facility, socio-cultural facility, entertainment facility, health facility, transportation, land use, economic condition and financial unit. Santoso (2000) use regression model to predict percentage of poor family at West Java villages based on PODES 1996 that is

$$Y= 40.00 + 4.99\,X2 - 0.115\,X4 + 0.00850\,X8 + 0.466\,X9 + 3.45\,X13 - 6.18\,X14 + 4.26\,X15 + 0.835\,X17 - 0.078\,X19 - 1.65\,X21 + 0.707\,X22 + 2.84\,X23$$

where

Y = percentage of poor family
X2 = type of widest road (1 = asphalt , 0 = non-asphalt)
X4 = distance from nearest hospital (kilometer)
X8 = population density (person per km-square)
X9 = religious center ratio per 1000 person
X13 = trash disposal of majority household (1 = available, 0 = not)
X14 = Place of latrine of majority household (1 = private, 0 = public)
X15 = newspaper/magazine customer (1 = have, 0 = not)
X17 = percentage of household who able to finance their children/family to university
X19 = percentage of household have television
X21 = percentage of household have 4-wheeled vehicle
X22 =percentage of household have 2 or 3-wheeled vehicle
X23 = average number of family member

In this research we only investigate four major variable from regression equation above, i.e main road type (X2), existence of waste facility (X13), toilet type (X14) and existence of newspaper/magazine subscribers (X15). Because of characterization algorithms work on categorical value, if other numerical variable will be include they must be transform into categorical value first.

Processes on spatial data mining module shown at Figure 2.

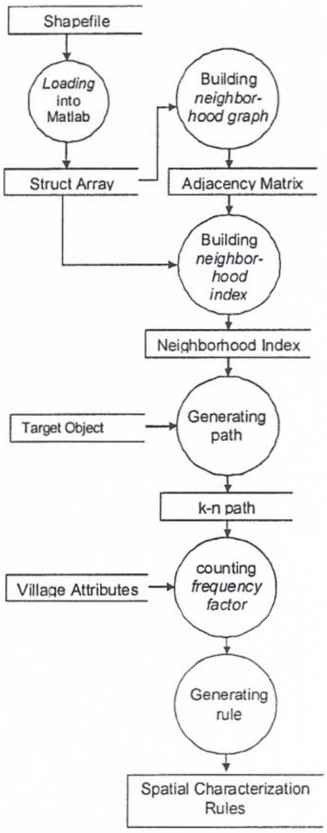


Figure 2. Processes on spatial data mining module

Using shaperead function, village shapefile is loaded into Matlab's environment and stored as struct array type variable. Figure 3 illustrated the structure of S variable.

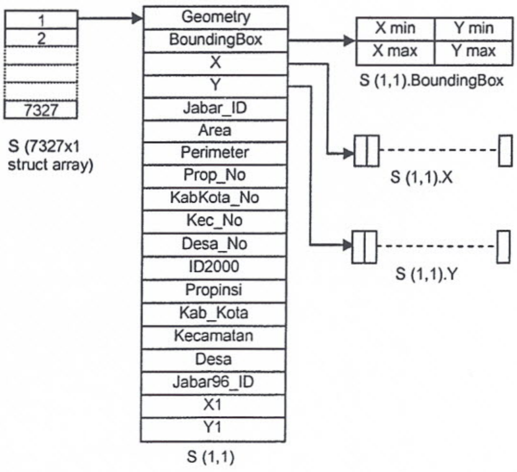


Figure 3. Struct array structure

The array shown at Figure 3 has number of element as much as number of polygon in the shapefile. Each element is a struct with following main fields

- Geometry : type of spatial object (point, line or polygon)
- BoundingBox : coordinates represents Minimum Bounding Rectangle i.e (x-min, y-min) and (x-max,y-max)
- X : x coordinates for points shapes a polygon.
- Y : x coordinates for points shapes a polygon
- X1 : x coordinate for polygon centre point
- Y1 : y coordinate for polygon centre point

Besides main fields above, the struct also have several fields like name and identifier for each village. In this prototype, each array element which is represent a village also called as object.

Our prototype are based on the concepts of neighbourhoods graphs and neighbourhood path which in turn are defined with respect to neighbourhood relations between objects (Ester *et al.* 1998). An object called as neighbour of another object if their boundaries meets at more than one node. Code to check neighbourhood between objects listed at Figure 4. Neighbourhood between objects stored in adjacency matrix and visualize with neighbourhood graph at Figure 5.

```
poligon2= M{poligon1}(j);
            poli1x = S(poligon1,1).X;
            poli1y = S(poligon1,1).Y;
            poli2x = S(poligon2,1).X;
            poli2y = S(poligon2,1).Y;
[xi,yi]=
polyxpoly(poli1x,poli1y,poli2x,poli2y,'unique');
adjacence = (size(xi,1)>1);
```

Figure 4. Code to check neighbourhood between objects

Polygon Number

	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	0	1	0	0	0	1
2	1	0	1	0	0	0	0	1	0	1
3	0	1	0	0	0	0	0	1	0	0
4	0	0	0	0	1	0	1	0	0	0
5	0	0	0	1	0	0	0	0	1	0
6	1	0	0	0	0	0	0	0	0	1
7	0	0	0	1	0	0	0	0	0	0
8	0	1	1	0	0	0	0	0	0	0
9	0	0	0	0	1	0	0	0	0	0
10	1	1	0	0	0	1	0	0	0	0

Polygon Number

Figure 5. Adjacency Matrix and Neighbourhood Graph

For each pair of adjacent object, we need to find distance and direction between them. The distance calculated from polygon's centre using Euclidian formula while direction found based on polygon's MBR and polygon's centre relative position (Figure 6).

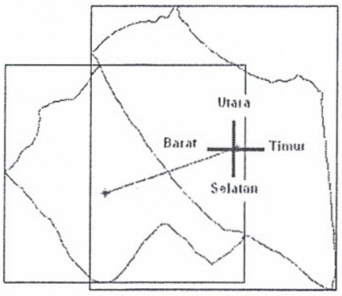
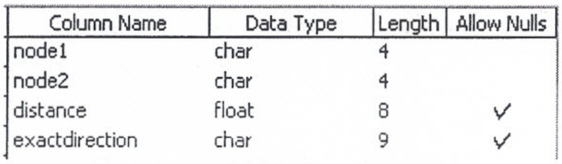


Figure 6. Distance and Direction

In order to avoid recalculation of distance and direction between objects in neighbourhood graph, we store distance and direction information into neighbourhood index. Because neighbourhood index has N by N size, its too large to stored at main memory at once. So we build neighbourhood index on SQL Server database management system. The structure of neighbourhood index table shown at Figure 7.



Column Name	Data Type	Length	Allow Nulls
node1	char	4	
node2	char	4	
distance	float	8	✓
exactdirection	char	9	✓

Figure 7. Neighbourhood Index Structure

Spatial characterization algorithm begin from the determination of target objects. Target object is a subset of the village identified as poor villages. From 177 villages in 8 districts of the sample was made by Santoso (2000), including 122 poor villages. From the poor villages, the 3 selected villages from each district set as the target object (Table 1). Figure 8 shows the distribution of the 24 villages targeted object (yellow) between the poor villages (red) and non-poor villages (green).

Table 1. Targeted Objects

NO	ID2000	PROPINSI	KAB_KOTA	KECAMATAN	DESA
1	3212180008	JAWA BARAT	INDRAMAYU	LOSARANG	PANGKALAN
2	3276050005	JAWA BARAT	KOTA DEPOK	BEJI	KUKUSAN
3	3276040009	JAWA BARAT	KOTA DEPOK	CIMANGGIS	HARJAMUKTI
4	3276020009	JAWA BARAT	KOTA DEPOK	PANCORAN MAS	PANCORAN MAS
5	3212120009	JAWA BARAT	INDRAMAYU	SLIYEG	GADINGAN
6	3201280008	JAWA BARAT	BOGOR	JASINGA	CURUG
7	3209230021	JAWA BARAT	CIREBON	GEGESIK	JAGAPURA KULON
8	3212010001	JAWA BARAT	INDRAMAYU	HAURGEULIS	BANTARWARU
9	3201130009	JAWA BARAT	BOGOR	SUKARAJA	CIUJUNG
10	3201160008	JAWA BARAT	BOGOR	CARIU	TANJUNG RASA
11	3209150008	JAWA BARAT	CIREBON	WERU	MEGUGEDE
12	3209040010	JAWA BARAT	CIREBON	BABAKAN	KARANGWANGUN
13	3204720005	JAWA BARAT	BANDUNG	CIMAHI TENGAH	PADASUKA
14	3204220001	JAWA BARAT	BANDUNG	GUNUNGHALU	CILANGARI
15	3205280006	JAWA BARAT	GARUT	KADUNGORA	KARANGTENGAH
16	3204150005	JAWA BARAT	BANDUNG	ARJASARI	LEBAKWANGI
17	3205180008	JAWA BARAT	GARUT	TAROGONG	JAYARAGA
18	3205030010	JAWA BARAT	GARUT	BUNGBULANG	GUNAMEKAR
19	3604190005	BANTEN	SERANG	MANCAK	BALEKAMBANG
20	3604150008	BANTEN	SERANG	CIPOCOK JAYA	PANANCANGAN
21	3604010009	BANTEN	SERANG	CINANGKA	RANCASANGGAL
22	3601120006	BANTEN	PANDEGLANG	LABUAN	TELUK
23	3601170004	BANTEN	PANDEGLANG	BANJAR	CIBEUREUM
24	3601050002	BANTEN	PANDEGLANG	CIGEULIS	KARANGBOLONG

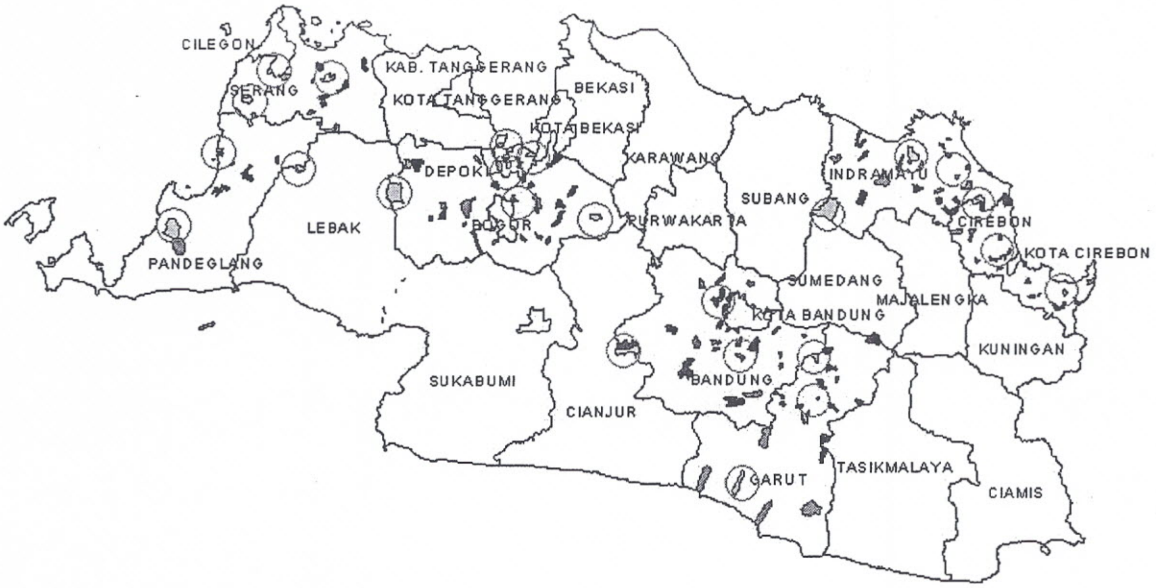


Figure 8. The distribution of the 24 villages targeted object (yellow) between the poor villages (red) and non-poor villages (green).

From target objects above, the algorithm need to create paths and extends to surrounding objects. Paths with two nodes (path k=2) represent direct neighbours of target objects. For path with larger number of nodes, we used filter predicate to determine the direction of path extensions. Path extensions are restricted such that satisfies each node only appears once in that path.

If a path only consist by two nodes, so those node called first-node and last-node. If a path have more than two nodes, so there is a before-last-node in that path. When a path is extended, then a filter predicate determine which node will be added to the path. Path extension process illustrated at Figure 9.

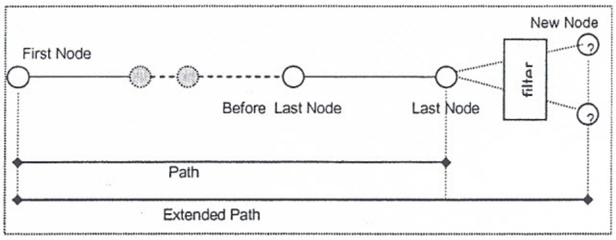


Figure 9. Path extension process.

We use filter starlike dan variable-starlike to extends a path. Ester *et al.* 2001 decribes those filter as follows. The filter starlike, e.g., is a very restrictive filter which allows only a small number of "coarse" paths. It requires that, when extending a path p = [n1,n2,...,nk] with a node nk+1, the exact "final" direction of p may not be generalized. For instance, a path with final direction northeast can only be extended by a node of an edge with the exact direction northeast. The filter variable-starlike allows more "finegrained" paths by requiring only that, when extending p the edge (nk, nk+1) has to fulfil at least the exact "initial" direction of p. For instance, a neighbourhood path with initial direction north can be xtended such that the direction north or the more special direction northeast is satisfied. Figure 10 shows those filter whereas codes concerned with that listed at Figure 11.

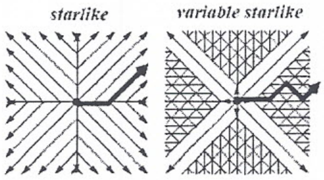


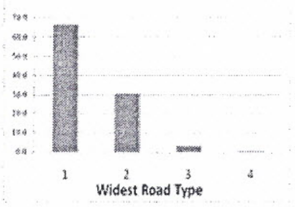
Figure 10. Filter Predicate

```
switch filter
case ('starlike')
[jarak_new, exact_direction_new]        = sdm_read_topology_db(db,lastnode,newnode);
[jarak_last, exact_direction_previous] = sdm_read_topology_db(db,before_lastnode,lastnode);
% absolute exact direction
if strcmp( exact_direction_new , exact_direction_previous);, tf=1;
  else
    if     strcmp(exact_direction_previous,'timur')& (strcmp(exact_direction_new,'tenggara') |
           strcmp(exact_direction_new,'timurlaut')), tf=1;
    elseif strcmp(exact_direction_previous,'barat') & (strcmp(exact_direction_new,'baratdaya') |
           strcmp(exact_direction_new,'baratlaut')), tf=1;
    elseif strcmp(exact_direction_previous,'selatan') & (strcmp(exact_direction_new,'tenggara')|
           strcmp(exact_direction_new ,'baratdaya')), tf=1;
    elseif strcmp(exact_direction_previous,'utara') & (strcmp(exact_direction_new,'baratlaut') |
           strcmp(exact_direction_new , 'timurlaut')), tf=1;
    else tf=0;
    end % end if
end %end if strcmp

case ('variable_starlike')
[jarak_new, exact_direction_new]        =  sdm_read_topology_db(db,firstnode,newnode);
[jarak_last, exact_direction_previous] =  sdm_read_topology_db(db,firstnode,lastnode);
if strcmp( exact_direction_new , exact_direction_previous);, tf=1;
  else % rel(i) is special relation of rel(1)
    if strcmp(exact_direction_new , 'tenggara') & ( strcmp(exact_direction_previous,'timur')|
       strcmp(exact_direction_previous,'selatan') ); tf=1;
    elseif strcmp(exact_direction_new , 'baratdaya') &(strcmp(exact_direction_previous,'barat')|
           strcmp(exact_direction_previous,'selatan') ); tf=1;
    elseif strcmp(exact_direction_new , 'baratlaut') &(strcmp(exact_direction_previous,'barat')|
           strcmp(exact_direction_previous,'utara') ); tf=1;
    elseif strcmp(exact_direction_new , 'timurlaut') &(strcmp(exact_direction_previous,'timur')|
           strcmp(exact_direction_previous,'utara') ); tf=1;
    else tf=0;
  end % end special relation
end % end check direction predicate
end % end switch filter

if tf==1;
extendedpath = [path newnode]; % extending path
counter = counter +1;
pathkn{counter,1}= extendedpath;
end %if tf
```

Figure 11. Path Extension Code

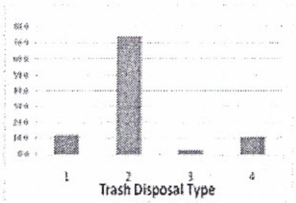
3. RESULT

Value distribution of West Java villages PODES2000 selected attributes shown at Figure 12 -15.



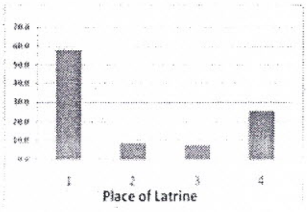
1-Asphalt/concrete/cone block
2-Hardened, (stone, pebble, etc)
3-Soil
4-Other

Figure 12. Type of widest road



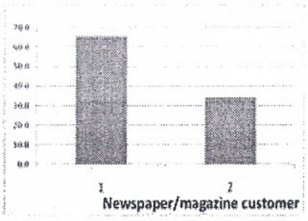
1-The trash then carried away
2-Buried in the hole
3-River
4-Otherwise

Figure 13. Trash disposal of majority household



1-Private toilet
2-Shared toilet
3-Public toilet
4-Not toilet

Figure 14. Place of latrine of majority household



1-Have
2-Not have

Figure 15. Newspaper/magazine customer

Seen from figures above, villages at West Java generally have good road, have newspaper/magazine customer and mostly of their household buried the trash in the hole and using private toilet. The *frequency factor* of prop = *selected attributes* with respect to *targets = poor villages* and *DB=all villages* shown at Figure 16 – 19 as numbers above the frequency bars.

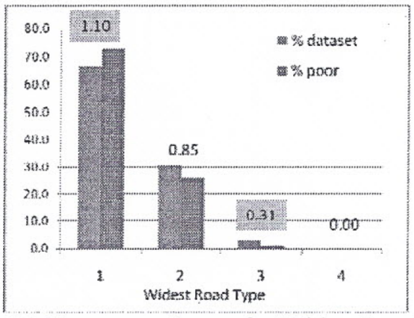


Figure 16. Frequency Factor for Widest Road Type
(*targets = poor villages* and *DB=all villages)*

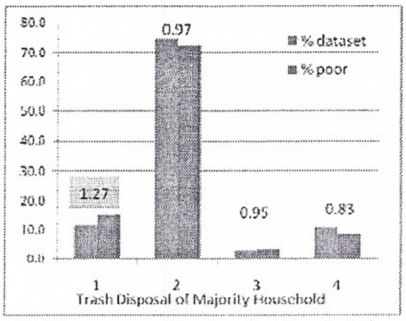


Figure 17. Frequency Factor for Trash Disposal of Majority Household
(*targets = poor villages* and *DB=all villages)*

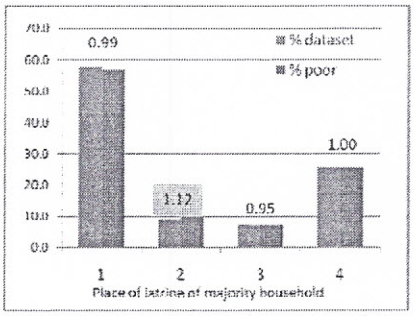


Figure 18. Frequency Factor for Place of Latrine of Majority Household
(*targets = poor villages* and *DB=all villages)*

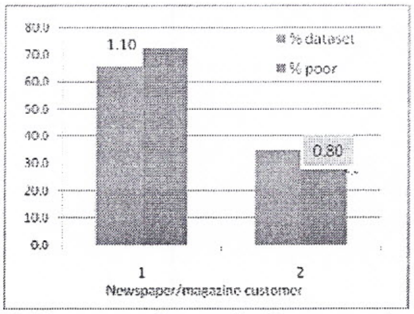


Figure 19. Frequency Factor for Newspaper/magazine customer
(*targets = poor villages* and *DB=all villages)*

High frequency factor for each attribute are highlighted. If we set *proportion* = 20 and *significance* = 1.25 (so 1/significance = 0.80) only one attribute value which characterize poor village i.e. Not Have Newspaper/Magazine Customers. In fact attribute value Soil as Widest Road Type have biggest significance value, but it can't characterize poor value because its occurrence is very low.

Next, the characterization algorithms will create direct path form target object, called k=2 path (Figure 20). Furthermore it extend "far away" from target objects consists of 3 nodes called k=3 path. Based on filter predicate used we have k=3-starlike path (Figure 21) and k=3-variable-starlike path (Figure 22). Path extended only till=3, because in that condition many paths getting through sub-district (kecamatan) boundaries.

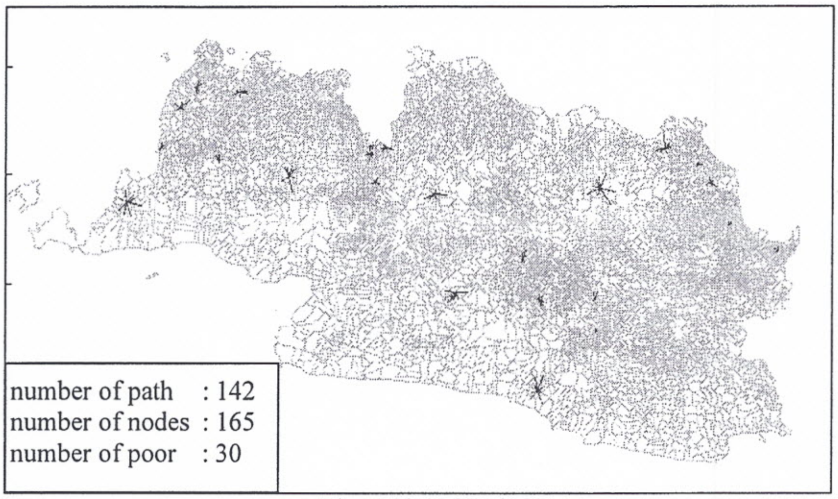


Figure 20. k=2 Path

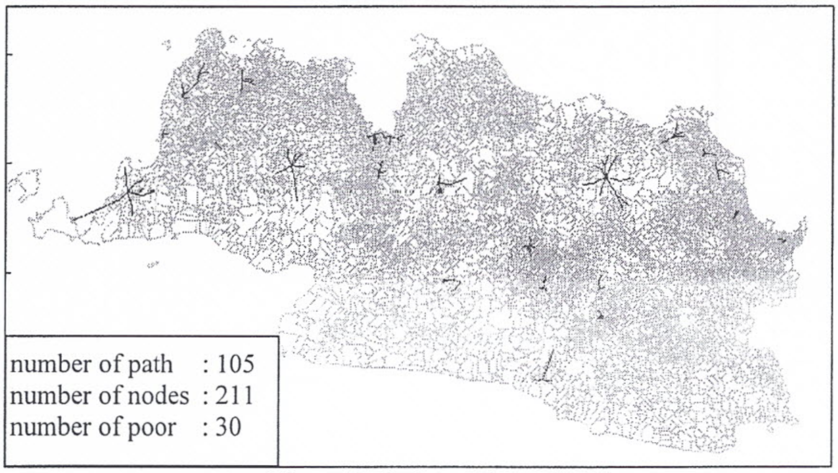


Figure 21. k=3-starlike path

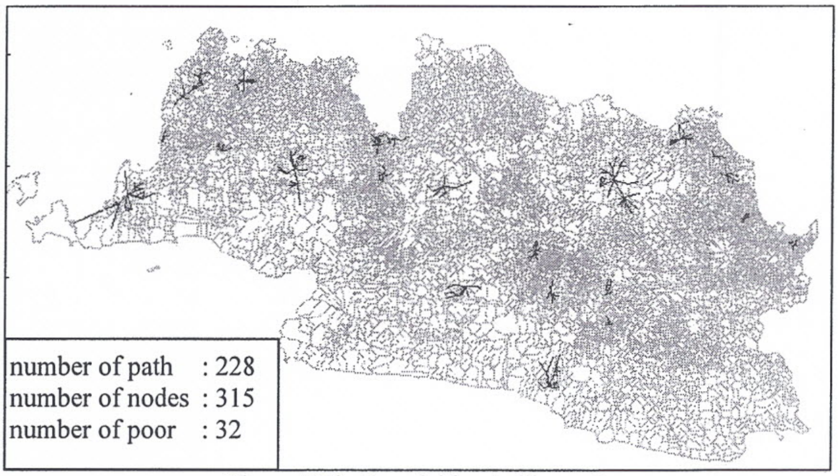


Figure 22.k=3-variable-starlike path

k=2-path has smallest number of node because it act as path extension initial point. Smallest number of path occurs at k=3-starlike-path. From 142 paths with k=2, only 105 paths which has extension using starlike filter, that shows starlike filter is very rigid. Variable-starlike filter seems very flexible since it has largest number of nodes and paths.

We don't have complete information about village poverty status at study area, so assume other villages except 122 poor villages determined by Santoso (2000) as non-poor villages. With that assumption we get all of paths above has number of poor nodes around 30. Frequency factor for those path shown at Table 1-4.

Table 1. Frequency Factor for Type of Widest Road Attribute

Type of widest road

Value	Label	dataset	% dataset	poor	% poor	freq-factor
1	Asphalt/concrete/cone block	4148	66.4	82	73.2	1.10
2	Hardened, (stone, pebble, etc)	1904	30.5	29	25.9	0.85
3	Soil	182	2.9	1	0.9	0.31
4	Other	9	0.1	0	0.0	0.00
		6243	100	112	100	

K3 Starlike	% K3 Starlike	K3 starlike poor	% K3 starlike poor	freq-factor
136	73.1	23	82.14	1.12
46	24.7	5	17.86	0.72
3	1.6	0	0.00	0.00
1	0.5	0	0.00	0.00
186	100	28	100	

Value	Label	K2	% K2	K2-poor	% K2-poor	freq-factor
1	Asphalt/concrete/cone block	104	71.2	23	82.1	1.15
2	Hardened, (stone, pebble, etc)	38	26.0	5	17.9	0.69
3	Soil	3	2.1	0	0.0	0.00
4	Other	1	0.7	0	0.0	0.00
		146	100	28	100	

K3 Varstarlike	% K3 Varstarlike	K3 varstarlike poor	%K3 varstarlike poor	freq-factor
191	68.71	25	83.33	1.21
80	28.78	5	16.67	0.58
6	2.16	0	0.00	0.00
1	0.36	0	0.00	0.00
278	100	30	100	

On all paths, frequency-factor for Value 1 (asphalt/cone/cone block) are greater than one with proportion at least 73 percent. This means road condition is good even for poor villages – the facts that only 2.9 % villages from the dataset are still using soil road. On the contrary, frequency-factor for Value 2 (hardened, stone, pebble, etc) are less than one for all paths with proportion not more than 25%. That's means the occurrences of hardened road are fewer in poor village than its surrounding area. The strongest characterization rule generated is

poor → type of widest road is asphalt/concrete/coneblock
(frequency-factor = 1.21, proportion 83.33)

Table 2. Frequency Factor for Trash Disposal of Major Attribute

Trash disposal of majority household

Value	Label	dataset	% dataset	poor	% poor	freq-factor
1	The trash then carried away	744	11.9	17	15.2	1.27
2	Buried in the hole	4656	74.5	81	72.3	0.97
3	River	172	2.8	4	3.6	1.30
4	Otherwise	676	10.8	10	8.9	0.83
		6248	100	112	100	

K3 Starlike	% K3 Starlike	K3 starlike poor	% K3 starlike poor	freq-factor
24	12.9	4	14.29	1.11
136	73.1	19	67.86	0.93
4	2.2	2	7.14	3.32
22	11.8	3	10.71	0.91
186	100	28	100	

Value	Label	K2	% K2	K2-poor	% K2-poor	freq-factor
1	The trash then carried away	22	15.1	4	14.3	0.95
2	Buried in the hole	96	65.8	18	64.3	0.98
3	River	4	2.7	2	7.1	2.61
4	Otherwise	24	16.4	4	14.3	0.87
		146	100	28	100	

K3 Varstarlike	% K3 Varstarlike	K3 varstarlike poor	%K3 varstarlike poor	freq-factor
36	12.95	4	13.33	1.03
191	68.71	21	70.00	1.02
8	2.88	2	6.67	2.32
43	15.47	3	10.00	0.65
278	100	30	100	

Table 2 shows there exist very small portion of poor village (max 7%) with majority of their household dispose the trash in the river. That condition expressed as rule below

poor →the trash dispose in the river
(frequency-factor = 3.32, proportion 7.14)

Table 3. Frequency Factor for Place of Latrine of Majority Household Attribute

Place of latrine of majority household

Value	Label	dataset	% dataset	poor	% poor	freq-factor
1	Private toilet	3615	57.9	64	57.1	0.99
2	Shared toilet	546	8.7	11	9.8	1.12
3	Public toilet	471	7.5	8	7.1	0.95
4	Not toilet	1616	25.9	29	25.9	1.00
		6248	100	112	100	

K3 Starlike	% K3 Starlike	K3 starlike poor	% K3 starlike poor	freq-factor
91	48.9	15	53.57	1.09
20	10.8	4	14.29	1.33
10	5.4	1	3.57	0.66
65	34.9	8	28.57	0.82
186	100	28	100	

Value	Label	K2	% K2	K2-poor	% K2-poor	freq-factor
1	Private toilet	65	44.5	15	53.6	1.20
2	Shared toilet	16	11.0	4	14.3	1.30
3	Public toilet	6	4.1	1	3.6	0.87
4	Not toilet	59	40.4	8	28.6	0.71
		146	100	28	100	

K3 Varstarlike	% K3 Varstarlike	K3 varstarlike poor	%K3 varstarlike poor	freq-factor
132	47.48	17	56.67	1.19
24	8.63	4	13.33	1.54
11	3.96	1	3.33	0.84
111	39.93	8	26.67	0.67
278	100	30	100	

In all paths generated, more than half of poor villages use private toilet as majority of its household place of latrine. That condition is not different in comparison with the other villages, average frequency factor is 1,1. Greatest frequency-factor occurs at K3-varstarlike path for shared-toilet attribute value, expressed as rule below

Poor → shared-toilet
(frequency-factor = 1.54, proportion = 13.33)

Table 4. Frequency Factor for Newspaper/Magazine Customer Attribute

Newspaper/magazine customer

Value	Label	dataset	% dataset	miskin	% poor	freq-factor
1	Have	4098	65.6	81	72.3	1.10
2	Not have	2150	34.4	31	27.7	0.80
		6248	100	112	100	

K3 Starlike	% K3 Starlike	K3 starlike poor	% K3 starlike poor	freq-factor
121	65.1	19	67.86	1.04
65	34.9	9	32.14	0.92
186	100	28	100	

Value	Label	K2	% K2	K2-poor	% K2-poor	freq-factor
1	Have	99	67.8	20	71.4	1.05
2	Not have	47	32.2	8	28.6	0.89
		146	100	28	100	

K3 Varstarlike	% K3 Varstarlike	K3 varstarlike poor	%K3 varstarlike poor	freq-factor
180	64.75	22	73.33	1.13
98	35.25	8	26.67	0.76
278	100	30	100	

Table 4 shows approximately 65 percent villages have newspaper/magazine customer, even higher (70 percent) for poor villages. Though this fact is weak because its frequency-factor not so much bigger than one, it may need to explore further rule below.

poor → have newspaper/magazine customer (frequency-factor = 1.13, proportion = 73.33).

?

4. CONCLUSION

in this research has successfully built spatial data mining prototype to apply the characterization algorithm and executed on the data podes west java and banten year 2003. we have shown the steps in the algorithm were implemented to form characterization rules. because of limited data that is used so that the characterization rules need to be further analysis. while the modules in the prototype can be used for other spatial data mining research.

REFERENCES

Ester M, Frommelt A, Kriegel H, Sander J. 1998. Algorithms for Characterization and Trend Detection in Spatial Databases. In : *Proceeding 4th International Conference on Knowledge Discovery and Data Mining* (KDD-98).

Ester M. Kriegel H, Sander. 2001. Algorithms and Applications for Spatial Data Mining. In : *Geographic Data Mining and Knowledge Discovery, Research Monographs in GIS.* Taylor and Francis.

Santoso A, 2000. Criteria of the Village Poor Families Based on the Calorie Consumption and its Relationship with Village Potential Statistics. [thesis]. Bogor: Institut Pertanian Bogor. Postgraduate Program.

ACKNOWLEDGEMENTS

The author grateful to Dr. Agus Buono (Department of Computer Science IPB) and Dr. Baba Barus (Departement of Soil Science and Land Resources IPB) who has give a lot of input in the implementation of this research

?