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

H.A. Adrianto

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 "retirees 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 $G_{\text{neighbour}}^DB$ be a neighbourhood graph and targets be a subset of $DB$. Let $freq(prop)$ 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

$$f_{\text{rel}}(prop) = \frac{freq(\text{targets}_s) \cdot freq(\text{DB}_s)}{card(\text{targets}) \cdot card(DB)}$$

Let significance and proportion be real numbers and let max-neighbours be a natural number. Let Neighbours $\text{Neighbours}_G^s(t)$ 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 = $\text{Neighbours}_G^s(\text{targets})$ as well as
2. the set objects = $\text{Neighbours}_G^n(t)$

for at least proportion many $t \in \text{targets}$ satisfy the condition:

$$f_{\text{object}}(prop) \geq \text{significance} \quad \text{or} \quad \frac{1}{\text{significance}}$$

The generated rule has the following format:

$$\text{target} \Rightarrow p_1(n_1, \text{freq-fac}_1) \land ... \land p_k(n_k, \text{freq-fac}_k).$$
This rule means that for the set of all targets extended by \( n \) neighbours, the property \( p \) is \( \text{freq-fac} \) times more (or less) frequent than in the database. So that the general spatial characterization algorithm works as follows (Ester et al. 2001):

\[
\text{discover-spatial-characterization (graph } G_{\text{DB}}, \text{ set of objects } \text{target}; \text{ real significance, proportion}; \text{ integer } \text{max-neighbours})
\]

1. Initialize the set of characterizations as empty;
2. Initialize the set of regions to targets;
3. Initialize \( n \) to 0;
4. Calculate \( \text{frequency}_{\text{DB}}(\text{prop}) \) for all \( \text{prop} = (\text{attribute}, \text{value}) \) in DB;
5. While \( n \leq \text{max-neighbours} \) do:
   a. For each attribute of DB and for the special attribute object type do:
      i. For each value of attribute do:
         1. Calculate \( \text{frequency}^{\text{region}}(\text{prop}) \) for property \( \text{prop} = (\text{attribute}, \text{value}) \);
         2. If \( f_{\text{region}}(\text{prop}) \geq \text{significance} \) or \( f_{\text{region}}(\text{prop}) \leq 1/\text{significance} \) then:
            i. Add \( (\text{prop}, n, f_{\text{region}}(\text{prop})) \) to the set characterizations;
         3. If \( n < \text{max-neighbours} \) then:
            i. For each object in region do:
               1. Add neighbours \( (G_{\text{DB}}, \text{object}, \text{TRUE}) \) to region;
               2. Increment \( n \) by 1;
         4. Extract all tuples \( (\text{prop}, n, f_{\text{region}}(\text{prop})) \) from characterizations which are significant in at least proportion of the regions with \( n \) extensions;
   b. 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.

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

\[
Y = 40.00 + 4.99X_2 - 0.115X_4 + 0.00850X_8 + 0.466X_9 + 3.45X_{13} - 6.18X_{14} + 4.26X_{15} + 0.835X_{17} - 0.078X_{19} - 1.65X_{21} + 0.707X_{22} + 2.84X_{23}
\]

where

\[
\begin{align*}
Y &= \text{percentage of poor family} \\
X_2 &= \text{type of widest road (1 = asphalt, 0 = non-asphalt)} \\
X_4 &= \text{distance from nearest hospital (kilometer)} \\
X_8 &= \text{population density (person per km-square)} \\
X_9 &= \text{religious center ratio per 1000 person} \\
X_{13} &= \text{trash disposal of majority household} \\
&\quad (1 = \text{available}, 0 = \text{not}) \\
X_{14} &= \text{Place of latrine of majority household} \\
&\quad (1 = \text{private}, 0 = \text{public}) \\
X_{15} &= \text{newspaper/magazine customer} (1 = \text{have}, 0 = \text{not}) \\
X_{17} &= \text{percentage of household who able to finance their children/family to university} \\
X_{19} &= \text{percentage of household have television} \\
X_{21} &= \text{percentage of household have 4-wheeled vehicle} \\
X_{22} &= \text{percentage of household have 2 or 3-wheeled vehicle} \\
X_{23} &= \text{average number of family member}
\end{align*}
\]

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.

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.

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 : y coordinates for points shapes a polygon
- XI : 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.
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.

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Data Type</th>
<th>Length</th>
<th>Allow Nulls</th>
</tr>
</thead>
<tbody>
<tr>
<td>node1</td>
<td>char</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>node2</td>
<td>char</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>distance</td>
<td>float</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>exactdirection</td>
<td>char</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>NO</th>
<th>ID2000</th>
<th>PROPINSI</th>
<th>KAB KOTA</th>
<th>KECAMATAN</th>
<th>DESA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3212710001</td>
<td>JAWA BARAT</td>
<td>BANDUNG</td>
<td>SUBANG</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3276050006</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>SUBANG</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3170400009</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>MUNCI</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3212707007</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>CINDAYA</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3276143003</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>SUDANG</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3201126008</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>KADARAN</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3209259002</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>TIRTA</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>3212714009</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>JAYA</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>3276113009</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>JAYA</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>3201126009</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>JAYA</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>3209124010</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>JAYA</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>3212710001</td>
<td>JAWA BARAT</td>
<td>BANDUNG</td>
<td>SUBANG</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>3276050006</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>SUBANG</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>3209259002</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>TIRTA</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>3212714009</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>JAYA</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>3201126008</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>KADARAN</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>3209124010</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>JAYA</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>3212710001</td>
<td>JAWA BARAT</td>
<td>BANDUNG</td>
<td>SUBANG</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>3276050006</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>SUBANG</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>3209259002</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>TIRTA</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>3212714009</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>JAYA</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>3201126008</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>KADARAN</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>3209124010</td>
<td>JAWA BARAT</td>
<td>KOTA BANDUNG</td>
<td>JAYA</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>3212710001</td>
<td>JAWA BARAT</td>
<td>BANDUNG</td>
<td>SUBANG</td>
<td></td>
</tr>
</tbody>
</table>

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 describes 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 \(p = [n_1, n_2, ..., n_k]\) with a node \(n_{k+1}\), 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 (\(n_k, n_{k+1}\)) has to fulfill at least the exact “initial” direction of \(p\). For instance, a neighbourhood path with initial direction north can be extended 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);
    if strcmp(exact_direction_previous, 'timur') || (strcmp(exact_direction_new, 'tínggara') ||
      strcmp(exact_direction_new, 'timurlaut')) tf=1;
    elseif strcmp(exact_direction_previous, 'barat') || (strcmp(exact_direction_new, 'baratdaya') ||
      strcmp(exact_direction_new, 'barat laut')) tf=1;
    elseif strcmp(exact_direction_previous, 'utara') || (strcmp(exact_direction_new, 'utara') ||
      strcmp(exact_direction_new, 'timurlaut')) tf=1;
    else t=0;
    end
  end
  % end switch
end
```

```
switch filter
  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, 'timur') || (strcmp(exact_direction_new, 'tenggara') ||
      strcmp(exact_direction_new, 'timurlaut')) tf=1;
    elseif strcmp(exact_direction_new, 'barat') || (strcmp(exact_direction_new, 'baratdaya') ||
      strcmp(exact_direction_new, 'barat laut')) tf=1;
    elseif strcmp(exact_direction_new, 'utara') || (strcmp(exact_direction_new, 'utara') ||
      strcmp(exact_direction_new, 'timurlaut')) tf=1;
    else tf=0;
    end
  end
end
```

Figure 11. Path Extension Code
3. RESULT

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

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

Figure 12. Type of widest road

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

Figure 13. Trash disposal of majority household

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

Figure 14. Place of latrine of majority household

I-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.

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.
k=2-path has smallest number of node because it acts 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

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
<th>dataset</th>
<th>% dataset</th>
<th>poor</th>
<th>% poor</th>
<th>freq.factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Asphalt/concrete/cone block</td>
<td>414</td>
<td>66.4</td>
<td>62</td>
<td>73.2</td>
<td>1.10</td>
</tr>
<tr>
<td>2</td>
<td>Hardened, (stone, pebble, etc)</td>
<td>1904</td>
<td>30.5</td>
<td>29</td>
<td>29.9</td>
<td>0.95</td>
</tr>
<tr>
<td>3</td>
<td>Soil</td>
<td>187</td>
<td>2.8</td>
<td>1</td>
<td>6.6</td>
<td>0.31</td>
</tr>
<tr>
<td>4</td>
<td>Other</td>
<td>9</td>
<td>0.1</td>
<td>0</td>
<td>0.0</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 2. Frequency Factor for Trash Disposal of Major Attribute

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
<th>dataset</th>
<th>% dataset</th>
<th>poor</th>
<th>% poor</th>
<th>freq.factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The trash then carried away</td>
<td>744</td>
<td>11.9</td>
<td>17</td>
<td>15.2</td>
<td>1.27</td>
</tr>
<tr>
<td>2</td>
<td>Buried in the hole</td>
<td>656</td>
<td>74.5</td>
<td>81</td>
<td>72.3</td>
<td>0.97</td>
</tr>
<tr>
<td>3</td>
<td>River</td>
<td>172</td>
<td>2.8</td>
<td>4</td>
<td>3.6</td>
<td>1.30</td>
</tr>
<tr>
<td>4</td>
<td>Otherwise</td>
<td>616</td>
<td>10.3</td>
<td>10</td>
<td>9.9</td>
<td>0.83</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Value</th>
<th>Label</th>
<th>dataset</th>
<th>% dataset</th>
<th>poor</th>
<th>% poor</th>
<th>freq.factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Private toilet</td>
<td>361</td>
<td>57.9</td>
<td>64</td>
<td>57.1</td>
<td>0.99</td>
</tr>
<tr>
<td>2</td>
<td>Shared toilet</td>
<td>541</td>
<td>87.7</td>
<td>11</td>
<td>9.6</td>
<td>1.12</td>
</tr>
<tr>
<td>3</td>
<td>Public toilet</td>
<td>471</td>
<td>7.5</td>
<td>8</td>
<td>7.1</td>
<td>0.95</td>
</tr>
<tr>
<td>4</td>
<td>Not toilet</td>
<td>1608</td>
<td>25.6</td>
<td>29</td>
<td>25.9</td>
<td>1.00</td>
</tr>
</tbody>
</table>

On all paths, frequency-factor for Value 1 (asphalt/concrete/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 means the occurrences of hardened road are fewer in poor village than its surrounding area. The strongest characterization rule generated is:

**poor \(\rightarrow\) type of widest road is asphalt/concrete/coneblock**

(frequency-factor = 1.21, proportion 83.33)

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 \(\rightarrow\) the trash dispose in the river**

(frequency-factor = 3.32, proportion 7.14)

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 \(\rightarrow\) shared-toilet**

(frequency-factor = 1.54, proportion 13.33)
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

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


ACKNOWLEDGEMENTS

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