Data mining for clustering naming of the village at Java Island

Atje Setiawan Abdullah¹, Budi Nurani Ruchjana², Akik Hidayat¹, Akmal¹, and Deni Setiana¹

¹ Department of Computer Science, Faculty of Mathematics and Natural Sciences
Universitas Padjadjaran, Indonesia
² Department of Mathematics, Faculty of Mathematics and Natural Sciences
Universitas Padjadjaran, Indonesia
E-mail: atjesetiawan@gmail.com

Abstract. Clustering of query based data mining to identify the meaning of the naming of the village in Java island, done by exploring the database village with three categories namely: prefix in the naming of the village, syllables contained in the naming of the village, and full word naming of the village which is actually used. While syllables contained in the naming of the village are classified by the behaviour of the culture and character of each province that describes the business, feelings, circumstances, places, nature, respect, plants, fruits, and animals. Sources of data used for the clustering of the naming of the village on the island of Java was obtained from Geospatial Information Agency (BIG) in the form of a complete village name data with the coordinates in six provinces in Java, which is arranged in a hierarchy of provinces, districts / cities, districts and villages. The research method using KDD (Knowledge Discovery in Database) through the process of preprocessing, data mining and postprocessing to obtain knowledge. In this study, data mining applications to facilitate the search query based on the name of the village, using Java software. While the contours of a map is processed using ArcGIS software. The results of the research can give recommendations to stakeholders such as the Department of Tourism to describe the meaning of the classification of naming the village according to the character in each province at Java island.

1. Introduction

How the ancestors in Java provides the naming of the village? Is there a difference in giving naming villages in each province in Java (West Java, Banten, Jakarta, Central Java, East Java, DIY)? How can we explain the naming of most villages by the prefix, the syllables are contained, and complete words, for every province in Java, and how its contribution to the naming of the village in Indonesia? This problem can be solved using ethnoscience approach [3].

Ethnoscience studied at culture with a scientific perspective. Ethnoscience helps to understand how people develop with different forms of knowledge and beliefs, and focuses on the ecological and historical contributions people have been given [4]. Ethnoscience is a new term and study that came into anthropological theory in the 1960s. Often referred to as “indigenous knowledge,” Ethnoscience introduces a perspective based on native perceptions. It is based on a complete emic perspective, which excludes all observations, interpretations and or any personal notions belonging to the ethnographer.
The taxonomy and clustering of indigenous systems, to name a few, used to categorize plants, animals, religion and life is adapted from a linguistic analysis. The concept of Native Science is also related to the understanding the role of the environment intertwined with the meaning humans place upon their lives. Understanding the language and the native peoples linguistic system is one method to understand a native peoples system of knowledge of organization.

In the year 2016, the Ministry of Tourism Indonesia again proposed a Pesona Indonesia or Wonderful Indonesia as a country branding and slogan tourist destinations in the country [5]. As we mention before the name of villages in Indonesia, is one of interested topic to be researched, because we can use multidicipline of sciences to learn about the pattern of villages naming based on large database for many purposes. Not only for tourism, but also for supporting a toponym as the general name for any place or geographical entity. Related, more specific types of toponym include hydronym for a body of water and oronym for a mountain or hill. Place names provide the most useful geographical reference system in the world. A toponymist relies not only on maps and local histories, but interviews with local residents to determine names with established local usage. Usually the naming village can be categorized by lexical, mythos, eventual, leadership and religion.

In this research we propose the spatial data mining using Knowledge Discovery in Database (KDD) method for mapping and description of village’s naming database in Java island at Indonesia via ethno-informatics approach [2]. Ethno-informatics is the discipline which it has the objective to study the basic law of information activities of culture supported by information technology. To get the good result in research of ethno-informatics, we should have relationship with other discipline, such as history, art, computer science, sociology, culture also statistics.

The special objectives of this research of visualization of villages naming for supporting tourism and indigenous culture in Indonesia can be noticed following:

(i) To get clustering of database villages naming in Java island using structure:
   - prefix naming of the village
   - initial syllable of villages naming
   - syllables contained in the naming of the village
   - complete word naming the village

In order to obtain the highest village naming information, both on the island of Java, as well as in each of the provinces in Java, West Java, Center of Java, East Java, DIY, Banten and DKI Jakarta.

(ii) To cluster database of villages naming of the village at Java island based on 16 categories that implies the following: the place, the environment, nature, feelings, herbs, vegetables, animals, poultry, fish, fruits, respect, state, color, creed, directions, and business.

(iii) To visualize database naming the village on the island of Java into map locations, to obtain picture of the position group or naming a particular village on a map of the island of Java.

(iv) To illustrate the database of villages naming at Java island in Venn diagram, so it can be shown:
   - The names of villages that characterizes the island of Java and in each province in Java
   - Intersection or influence the naming of the village between the provinces in Java, so it can be known to the provinces where the percentage rate of the high similarit
   - Developing theorems or mathematical models associated with intersection of nearest neighbors at each province in Java island.

(v) To develop an application program based on query for villages naming based query.

(vi) To develop a data mining application for the villages naming which integrated manner, in order to facilitate the user in understanding the meaning of the villages naming of the village on the island of Java and in each province in Java.
2. Research Method

The method of this research based on KDD method. The first step, we build of the study design, utilizing the science of the latter being developed, known as data mining, which is defined to extract knowledge automatically from large databases, to obtain the patterns interesting, thus forming a knowledge [6]. The KDD method can be described as follow:

Data mining process includes data preprocessing, data mining, and postprocessing. Data preprocessing steps including a data cleaning, data transformation, data integration and data selection, the process of data mining including the use of models to process data, while the postprocessing including processing visualization and interpretation of results, and conclusion about the result to becomes a knowledge. In this research we used large database database village from Geographic Information Bureau of Indonesia and then we build a software to create information for village’s naming based on meaning of preposition, word and a whole word via descriptive statistics with supported by graphically and multivariate analysis. The research methodology for naming village can be built as Figure 2.
In this research, we built Java-based application program, to do the search results of the naming village at Java island.

3. Main Result
We have the recapitulation of database of village naming at Java island as following Table 1.

| NO | PROVINCES | NUMBER of VILLAGE | PRECENTAGE |
|----|------------|-------------------|------------|
| 1  | JABAR      | 5760              | 23.17      |
| 2  | BANTEN     | 1534              | 6.17       |
| 3  | DKI        | 267               | 1.07       |
| 4  | JATENG     | 8355              | 33.60      |
| 5  | JATIM      | 8509              | 34.22      |
| 6  | DIY        | 438               | 1.76       |
|    | TOTAL PULAU JAWA | 24863            | 100.00     |

Furthermore we built the software application for searching the whole word of village naming as Figure 3.

![Figure 3. Application Search the Whole Name of Village Naming](image-url)
In our application, we determine the prefix of village name at Java island to be Ci, Pa and Su and we got the result how many percentage of the name with three of prefix at each province at Java island. We explain the result by Figure 4.

![Figure 4. Example Sample Search for Prefix](image)

We can describe that the prefix Ci means water, Pa is a tool for to do something and Su is beautiful. But, sometimes the meaning is confuse, so we need to extend the prefix to be a word and a whole word. So, we can have the table of recapitulation the village naming at Java island as following Table II and III. We divide the Table to be two sub Table, Table II and Table III. We choose 16 word from the village name, with the meaning for the place, the environment, nature, feelings, herbs, vegetables, animals, poultry, fish, fruits, respect, state, color, creed, directions, and business. Table II consist of recapitulation 16 word at Provinces: Yogyakarta, Central of java and East Java. While the Table 3 including Provinces: West java, Jakarta, Total of Java Island and National for all of informaton at Indonesia.

| NO | WORDS          | LABEL           | DIY  | JATENG | JATIM |
|----|----------------|-----------------|------|--------|-------|
| 1  | SARI           | USAHA           | 26   | 593    | 375   |
| 2  | KARANG         | ALAM            | 16   | 379    | 213   |
| 3  | JAYA           | USAHA           | 1    | 24     | 18    |
| 4  | MULYO/MULYA    | PENGHORMATAN    | 21   | 163    | 110   |
| 5  | KALI           | ALAM            | 6    | 257    | 122   |
| 6  | JATI           | TUMBUHAN        | 6    | 143    | 115   |
| 7  | TIMUR/WETAN    | ARAH            | 1    | 124    | 151   |
| 8  | BARAT/KULON    | ARAH            | 1    | 114    | 151   |
| 9  | LOR/UTARA/KALER| ARAH            | 2    | 126    | 118   |
| 10 | AGUNG          | PENGHORMATAN    | 7    | 74     | 135   |
| 11 | TANJUNG        | ALAM            | 1    | 59     | 55    |
| 12 | GUNUNG         | ALAM            | 1    | 71     | 54    |
| 13 | TEGAL          | ALAM            | 4    | 71     | 47    |
| 14 | AWI/PRING/BAMBU| TUMBUHAN        | 5    | 50     | 31    |
Continue of Table III

Table 3. Recapitulation of Word in 16 Categories at Java Island

| NO WORDS | LABEL    | JABAR | BANTEN | DKI | NATIONAL | JAVA ISLAND |
|-----------|----------|-------|--------|-----|----------|-------------|
| 1 SARI USAHA | 546 | 25 | 4 | 2298 | 1569 |
| 2 KARANG ALAM | 157 | 20 | 1 | 1062 | 786 |
| 3 JAYA USAHA | 376 | 87 | 3 | 1543 | 509 |
| 4 MULYO/MUYA PENGHORMATAN | 178 | 19 | 1 | 881 | 492 |
| 5 KALI ALAM | 58 | 4 | 6 | 590 | 453 |
| 6 LATI TUMBUHAN | 126 | 9 | 7 | 492 | 406 |
| 7 TIMUR/WETAN ARAH | 81 | 24 | 12 | 742 | 393 |
| 8 BARAT/KULON ARAH | 76 | 21 | 12 | 720 | 375 |
| 9 LOR/UTARA/KALER ARAH | 80 | 11 | 19 | 678 | 356 |
| 10 AGUNG PENGHORMATAN | 16 | 13 | 3 | 569 | 248 |
| 11 TANJUNG ALAM | 82 | 13 | 4 | 1182 | 214 |
| 12 GUNUNG ALAM | 64 | 18 | 3 | 556 | 211 |
| 13 TEGAL ALAM | 42 | 12 | 2 | 225 | 178 |
| 14 AWI/PRING/BAMBU TUMBUHAN | 57 | 9 | 6 | 361 | 158 |

Furthermore, we gave the recapitulation of village naming in the whole word for each province.

Table 4. Village Naming at Each Province using A Full Name

| NO NAME OF VILLAGE | CLUSTERING | NATIONAL | JABAR | BANTEN | JATENG | DIY | DKI | JATIM | JAVA ISLAND | LABEL |
|---------------------|------------|----------|-------|--------|--------|-----|-----|-------|-------------|-------|
| 1 KARANGANYAR ALAM | 108 | 18 | 0 | 40 | 0 | 0 | 19 | 77 | JABAR,JATENG,JATIM |
| 2 BABAKAN TEMPAT | 71 | 51 | 9 | 7 | 0 | 0 | 1 | 68 | JABAR,BANTEN,JATENG,JATIM |
| 3 MUKURSI USAHA | 110 | 43 | 15 | 2 | 0 | 0 | 0 | 60 | JABAR,BANTEN,JATENG |
| 4 SIDOREO KEADAAN | 78 | 0 | 0 | 23 | 3 | 0 | 30 | 56 | JATENG,DIV,JATIM |
| 5 TANJUNGSAI ALAM | 83 | 25 | 2 | 16 | 0 | 0 | 11 | 54 | JABAR,BANTEN,JATENG,JATIM |
| 6 RANJARSARI TEMPAT | 67 | 8 | 6 | 21 | 1 | 0 | 18 | 54 | JABAR,BANTEN,JATENG,DIV,JATIM |
| 7 KABARANGSI ALAM | 64 | 14 | 1 | 24 | 2 | 0 | 6 | 47 | JABAR,BANTEN,JATENG,DIV,JATIM |
| 8 WONOBOYO PERASAAN | 60 | 0 | 0 | 16 | 0 | 0 | 20 | 45 | JATENG,JATIM |
| 9 MEKARSAH USAHA | 102 | 37 | 6 | 0 | 0 | 0 | 0 | 43 | JABAR,BANTEN |
| 10 REJONSI PERASAAN | 64 | 0 | 0 | 29 | 1 | 0 | 12 | 42 | JATENG,JATIM,JATIM |
| 11 KARANGSI ALAM | 57 | 0 | 1 | 18 | 0 | 0 | 23 | 40 | BANTEN,JATENG,JATIM |
| 12 SUKAMALUSA PENGHORMATAN | 65 | 35 | 3 | 0 | 0 | 0 | 0 | 38 | JABAR,BANTEN |
| 13 SUMBAHANG PENGHORMATAN | 70 | 0 | 0 | 10 | 2 | 0 | 26 | 38 | JATENG,JATIM,DIV |
| 14 UKUSAHI USAHA | 117 | 32 | 5 | 0 | 0 | 0 | 0 | 37 | JABAR,BANTEN |
| 15 GUNUNGSI ALAM | 63 | 11 | 3 | 9 | 1 | 0 | 11 | 35 | JABAR,BANTEN,JATENG,JATIM |
| 16 BABAKAN TEMPAT | 31 | 21 | 4 | 7 | 0 | 0 | 1 | 33 | JABAR,BANTEN,JATENG,DIV,JATIM |
| 17 JATSARI TUMBUHAN | 36 | 10 | 0 | 10 | 0 | 0 | 13 | 33 | JABAR,BANTEN,JATENG |
| 18 KABARANGSI PERASAAN | 54 | 0 | 0 | 14 | 0 | 0 | 10 | 33 | JATENG,DIV,JATIM |
| 19 SUKAMALUSA PERASAAN | 36 | 25 | 5 | 0 | 0 | 0 | 0 | 30 | JABAR,BANTEN |
| 20 NEGLASSI TEMPAT | 33 | 28 | 2 | 0 | 0 | 0 | 0 | 30 | JABAR,BANTEN |

Finally we also gave the comparison of how many data that we have for Recapitulation of a Full Name of naming village at West Java if we compare with the data in Indonesia at Table V.

Table 5. The Comparison of Full Name of Village Naming at West java Province and Indonesia (National)

| NO NAME OF VILLAGE | CLUSTERING | NATIONAL | JABAR | BANTEN | JATENG | DIY | DKI | JATIM | JAVA ISLAND | LABEL |
|---------------------|------------|----------|-------|--------|--------|-----|-----|-------|-------------|-------|
| 1 KARANGANYAR ALAM | 108 | 18 | 0 | 40 | 0 | 0 | 19 | 77 | JABAR,JATENG,JATIM |
| 2 BABAKAN TEMPAT | 71 | 51 | 9 | 7 | 0 | 0 | 1 | 68 | JABAR,BANTEN,JATENG,JATIM |
| 3 MUKURSI USAHA | 110 | 43 | 15 | 2 | 0 | 0 | 0 | 60 | JABAR,BANTEN,JATENG |
| 4 SIDOREO KEADAAN | 78 | 0 | 0 | 23 | 3 | 0 | 30 | 56 | JATENG,DIV,JATIM |
| 5 TANJUNGSAI ALAM | 83 | 25 | 2 | 16 | 0 | 0 | 11 | 54 | JABAR,BANTEN,JATENG,JATIM |
| 6 RANJARSARI TEMPAT | 67 | 8 | 6 | 21 | 1 | 0 | 18 | 54 | JABAR,BANTEN,JATENG,DIV,JATIM |
| 7 KABARANGSI ALAM | 64 | 14 | 1 | 24 | 2 | 0 | 6 | 47 | JABAR,BANTEN,JATENG,DIV,JATIM |
| 8 WONOBOYO PERASAAN | 60 | 0 | 0 | 16 | 0 | 0 | 20 | 45 | JATENG,JATIM |
| 9 MEKARSAH USAHA | 102 | 37 | 6 | 0 | 0 | 0 | 0 | 43 | JABAR,BANTEN |
| 10 REJONSI PERASAAN | 64 | 0 | 0 | 29 | 1 | 0 | 12 | 42 | JATENG,JATIM,JATIM |
| 11 KARANGSI ALAM | 57 | 0 | 1 | 18 | 0 | 0 | 23 | 40 | BANTEN,JATENG,JATIM |
| 12 SUKAMALUSA PENGHORMATAN | 65 | 35 | 3 | 0 | 0 | 0 | 0 | 38 | JABAR,BANTEN |
| 13 SUMBAHANG PENGHORMATAN | 70 | 0 | 0 | 10 | 2 | 0 | 26 | 38 | JATENG,JATIM,DIV |
| 14 UKUSAHI USAHA | 117 | 32 | 5 | 0 | 0 | 0 | 0 | 37 | JABAR,BANTEN |
| 15 GUNUNGSI ALAM | 63 | 11 | 3 | 9 | 1 | 0 | 11 | 35 | JABAR,BANTEN,JATENG,JATIM |
| 16 BABAKAN TEMPAT | 31 | 21 | 4 | 7 | 0 | 0 | 1 | 33 | JABAR,BANTEN,JATENG,DIV,JATIM |
| 17 JATSARI TUMBUHAN | 36 | 10 | 0 | 10 | 0 | 0 | 13 | 33 | JABAR,BANTEN,JATENG |
| 18 KABARANGSI PERASAAN | 54 | 0 | 0 | 14 | 0 | 0 | 10 | 33 | JATENG,DIV,JATIM |
| 19 SUKAMALUSA PERASAAN | 36 | 25 | 5 | 0 | 0 | 0 | 0 | 30 | JABAR,BANTEN |
| 20 NEGLASSI TEMPAT | 33 | 28 | 2 | 0 | 0 | 0 | 0 | 30 | JABAR,BANTEN |

Furthermore, we describe the result of village naming using map in Java island. So, we have the difference mapping of each criteria at Java island map. For example we can build the map of village naming of Karangsari a blue one and Wonorejo in a yellow. We can get the map that the village naming of Karangsari is at West Java-Center and also East Java, while the naming village of Wonorejo just only at Center and East Java.
We can continue to make a map of village naming of Mekarsari and Sukorejo and get the result as Figure 6. The map show us that the village naming of Mekarsari just only in West java, while the name of Sukorejo can be found at Center and East Java.

Furthermore to illustrate the similarity of database of villages naming at Java island we can use the Venn diagram. For example we can choose the similarity of village naming at West Jav-Center Java and East java as following Figure.

Using the Venn diagram above, we can describe the intersection of village naming between three provinces are: beringin, Bandar and Lingga. From this phenomena, finally we can continue
Figure 7. Venn Diagram of intersection of Villages Naming in Java Island

a mathematical work to get a theorem about the intersection of village naming in Java island.

4. Conclusion
(i) Village naming illustrate cultural behavior that is very important in daily life, past, present and future.
(ii) Community character Java can be described following criteria: has a soul strong and courageous, loving the country and the beauty of the environment, respecting local culture, having the ability of fisheries and livestock, ready to wander to another place, soulless strong and steadfast in the face of life, loves fruits local product, loved the beauty of flowers local products, can express the state under various conditions, thinking and acting well and enjoy the existing conditions, working hard to make a good situation, appreciate its leader, appreciate culture, behave religious, has a subtle feeling, love beauty, and well-behaved to anyone, loving home to live a life in peace, have a strong soul, use the plants as a supporting activity of life, be clean, and courageous, open and able to communicate with other cultures both locally and internationally.

Acknowledgment
The work reported in this paper is supported by the Academic Leadership Grant Unpad 2016, so the authors thank to Rector Universitas Padjadjaran and Directorate Research and Community Services Unpad for supporting funding of research and paper dissemination.

References
[1] Abdullah A S et al. 2015a Ethnomathematics in Cipatujah West Java, presented at LEAD Leiden Universiteit, the Netherlands
[2] Abdullah A S 2015b Data Mining for Villages Naming at West Java at National Seminar on Ethnomathematics at UHAMKA Jakarta
[3] Ahimsa-Putra H S 1985 Etnosains dan Etnometodologi: Sebuah Perbandingan dalam Masyarakat Indonesia Majalah Ilmu-ilmu Sosial Indonesia Jakarta: Lembaga Ilmu Pengetahuan Indonesia Jilid XII Nomor 2 pp 103–133
[4] Atran S 1991 Social Science Information / Sur Les Sciences Sociales. Ethnoscience Today 30(4) pp 595–662
[5] http://pesona.indonesia.travel/overview/detail/6/indonesia-sekilas
[6] Jiawei H, Kamber M, and Jian P 2012 Data Mining Concepts and Technique, Morgan Kaufmann Publishers, USA