



Analysis of Using Google Maps Data to Measure the Presence or Accessibility of Urban Facilities for The BPS - Statistics Indonesia's Classification of Urban and Rural Villages

Benny Firmansyah ^{1,2}, Widya Sri Wahyuni ^{3,*}

¹BPS - Statistics of West Sumatra Province, Indonesia

²School of Electrical Engineering and Informatics, Institut Teknologi Bandung

³BPS - Statistics of Bukittinggi Municipality, Indonesia

Email: ¹benny.firmansyah@bps.go.id, ²widyasw@bps.go.id

Abstract

The BPS - Statistics Indonesia classifies villages into urban villages and rural villages to organize statistics. The classification of village areas into urban or rural status is intended to form a stratum used in survey sampling techniques. With this status, it is hoped that the sample taken can represent the entire population well. The BPS - Statistics Indonesia establishes criteria for classifying a village into an urban village. The 2020 urban village criteria use three indicators as its measure, namely: population density per km², percentage of agricultural families, and the presence or access of urban facilities. In general, the data used in calculating the classification of urban and rural villages in 2020 uses data from the 2019 Village Potential (Podes) survey. This study utilizes data on urban facilities such as schools, markets, shops, and hospitals on the google maps website to calculate the score of indicators of the existence or access of urban facilities. This study used a web scraping method to obtain data on these urban facilities from the google maps website. This study selected eight villages in the Lubuk Sikaping District, Pasaman Regency, West Sumatra Province, as a case study. The results showed that four villages with great potential were classified into urban villages, and three villages with great potential were classified into rural villages.

Keywords: Classification of Villages, Urban Facilities, Google Maps, Web Scraping

1. INTRODUCTION

The territory of Indonesia is divided into several levels of administrative regions, namely provinces, regencies or cities, sub-districts, and villages. The village is the smallest administrative area and is often used as an observation unit. The villages covered include various villages with various legal statuses, such as traditional villages (both village and sub-district status), preparatory villages, Smallest Settlement Units (UPI), and Alienated Community Settlements (PMT) [1]. Each village has different socioeconomic characteristics of the community, features, and



environmental topology. These conditions will continue to change along with the level of development in a village. These changes in requirements are used as an indicator to classify a village into the classification of urban villages or rural villages.

To organize statistics, BPS - Statistics Indonesia classifies villages into urban and rural villages [1]. The classification by urban and rural areas is considered better to describe the characteristics of the village [1]. The category of village areas into urban or rural status is intended to form a stratum used in survey sampling techniques [2]. In addition, this classification is also used to allocate census officers or surveys to determine strategic estimate figures. With the status of urban and rural areas, it is hoped that the samples taken can represent the entire population well. In the analysis, classifying villages into urban or rural villages will provide results that better describe the actual situation compared to the classification of villages [2].

BPS - Statistics Indonesia sets the 2020 urban village criteria, consisting of three indicators as its measure: population density per km², percentage of agricultural families, and the existence of urban facilities or access to these urban facilities [1]. Each indicator provides a score that will determine the urban and rural villages. The data used in calculating the classification of rural-urban villages uses data from the 2019 Village Potential (Podes) survey. This research focuses on the hands of urban facilities or access to these urban facilities. Many urban facilities must be dated on the Podes survey, including kindergartens, junior high schools, public high schools, markets, shops, hospitals, hotels, billiards, pubs, discotheques, karaoke venues, and salons [1]. Data collection of various urban facilities requires a lot of costs, energy, and time.

Based on these problems, this study tried to survey the use of landmark data on various types of urban facilities on the google maps site. The landmark data was extracted using web scraping methods [3]–[6]. Furthermore, the data is processed to produce a score value for indicators of urban facilities' existence or access to these urban facilities. As a case study, eight villages were selected in Lubuk Sikaping District, Pasaman Regency, West Sumatra Province.

2. METHODS

This study focused on calculating scores for indicators of the presence or access of urban facilities through several types of landmark data extracted from google maps sites using web scraping methods. The landmark data is then overlaid with a digital map of village administration using QGIS software. Furthermore, the calculation of the score is carried out. The mindset of this study can be seen in Figure 1 below.

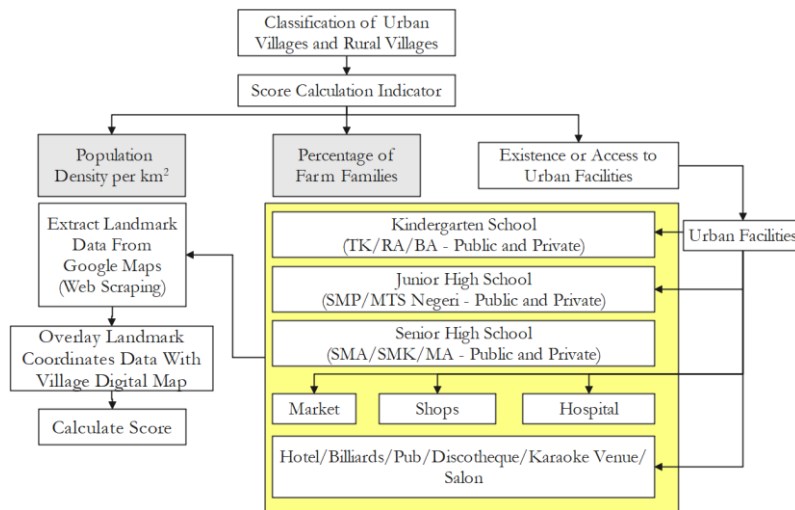


Figure 1. Research mindset

Some studies have utilized data extraction from google maps using web scraping methods for analysis or a particular study. Some of these studies can be seen in table 1 below.

Table 1. Some related studies that use google maps data extraction with web scraping method

Reference	Web Scraping Tool	Types of Extracted Data	Research Objectives
[6]	Botsol Crawler Software	Business data of food and beverage providers (Business name, address, postal code, coordinate point, and web address)	Collecting business data of food and beverage providers by web scraping methods, matching web scraping data with BPS-Statistics Indonesia data, and analyze areas that have potential business of food and beverage providers.
[7]	Selenium and Python	Traffic data (Direction from A to B and opposite)	Finding the most congested areas across the highway in both directions, time chart, the average time to travel in various hours such as morning, night, etc
[8]	Google Map's API	Populartime data records real-time human location or activities that are collected from users of Google's maps on smartphones	Monitoring electricity or energy demand by considering human activities in real-time.

2.1. Village Classification Score Calculation

Table 2. Variables, classifications, scores, and criteria of urban villages [1]

Criteria				Existence or access to urban facilities		
Population Density per km ²	Score	Percentage of Farm Families	Score	Urban Facilities	Criteria	Score
<500	1	> 70,00	1	Kindergarten School	Exist or < 2,5 km*	1
500 - 1.249	2	50, 00 - 69,99	2	Junior High School	> 2,5 km*	0
1.250 - 2.499	3	30, 00 - 49,99	3	Senior High School		
2.500 - 3.999	4	20,00 - 29,99	4	Market	Exist or < 2,5 km*	1
4.000 - 5.999	5	15,00 - 19,99	5	Shops	> 2,5 km*	0
6.000 - 7.499	6	10,00 - 14,99	6	Hospital	Exist or < 2,5 km*	1
7.500 - 8.499	7	5,00 - 9,99	7	Hotel/Billiards/Discotheque/ Pub/Karaoke Venue/Salon	> 2,5 km*	0
> 8.500	8	< 5,00	8		Exist	1
				Percentage of Families of Landline Users	None	0
				Percentage of Electricity User Families	≥ 2,00	1
					< 2,00	0
					≥ 95,00	1
					< 95,00	0

* Distance measured from the position of the village office

This study calculates scores for indicators of the presence or access to urban facilities. Table 2 above presents in full the variables, criteria, and scores used in the classification of urban villages and rural villages. Based on the requirements in Table 2, it is known that a village can achieve a maximum score of 25 and a minimum score of 2 [1]. The existence or access to urban facilities is also very influential, with a top score of 7 [1]. Meanwhile, the cut of points used to determine urban villages is 9 [1]. So villages with a score of 9 or more are designated as urban villages, while villages with a score of less than 9 are designated as rural villages [1].

2.2. Web Scraping

Web scraping is a technique for retrieving textual characters from web pages, primarily for analysis [9]. Research by [7] also defines web scraping as a web harvesting, a method that can be used to capture and extract a large amount of data from a website and store it in a structured format. The primary purpose of web scraping is to extract information from one or many websites and process it into simple structures such as spreadsheets, databases, or CSV (Comma Separated Values) files [10]–[13]. Many techniques have been used to retrieve content from a web page such as API computer languages, robots, intelligent agents, and web scraper software [14]. This study used Botsol Crawler, one type of web scraper software, to extract all the landmark data information of urban facilities from google maps sites. The data attributes taken the name of the landmark, the address, and the coordinate point of the landmark location. Data extraction through this software is done automatically using robots and without blocking [5].

3. RESULTS AND DISCUSSION

3.1. Landmark Extracts

The results of extracts of landmark information on urban facilities in Lubuk Sikaping District using Botsol software can be seen in Table 3. The scraping process with Botsol software can be seen in Figure 2. The attributes taken include landmark names, addresses, and coordinate points. Furthermore, the results of the landmark information extract are re-checked so that the subsequently processed landmarks are by the criteria for urban facilities in calculating the score for village classification, as previously explained in Table 2. In addition, an extract of information from the village office in Lubuk Sikaping District was also carried out as a basis for calculating the distance to the nearest urban facility if there were no certain urban facilities in the village.

Table 3. The results of extracting information on landmarks of urban facilities in Lubuk Sikaping District using Botsol software

Urban Facilities	Number of Landmarks After Checking
Kindergarten School	8
Junior High School	9
Senior High School	7
Market	5
Shops	1
Hospital	5
Hotel/Billiards/Discotheque/Pub/Karaoke Venue/Salon	26
Total	61

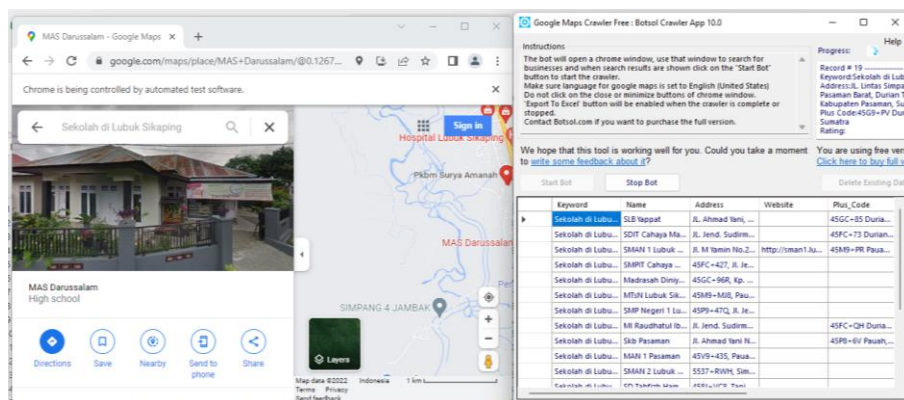


Figure 2. The process of scraping landmark information of urban facilities from google maps sites using Botsol software

3.2. Overlay Landmark Points of Urban Facilities With Digital Map of Villages

At this stage, an overlay of the coordinate points of the landmarks of urban facilities that have been extracted previously with a digital map of the village is carried out. The digital map of the village used is a map of the village administrative boundary from the Geospatial Information Agency (BIG). In comparison, the software used for overlays is QGIS. Previously, landmark point data in CSV format was converted into vector data in QGIS so that the overlay process could be carried out.

Based on the overlay results, the existence of urban facilities in certain village areas can be seen. Meanwhile, if there is none, then the distance of the nearest urban facility from the village office is calculated. The distance calculation uses the measure line feature of two landmark points in QGIS software. The process at this stage can calculate the score on the indicators of the presence or access of urban facilities in each village. In this study, we tried to calculate the score of the existence or access of urban facilities for eight villages in the Lubuk Sikaping District. Figure 3 and Figure 4 show an example of overlaying landmarks of public high schools and hotels with a digital map of village boundaries in Lubuk Sikaping District using QGIS software. Meanwhile, Figure 5 shows an example of calculating the distance of the nearest urban facility from the village office.



Figure 3. Overlay of landmark points of senior high schools with a digital map of village boundaries in Lubuk Sikaping District using QGIS software



Figure 4. Overlay of hotel landmark points with a digital map of the village boundary of Lubuk Sikaping District using QGIS software

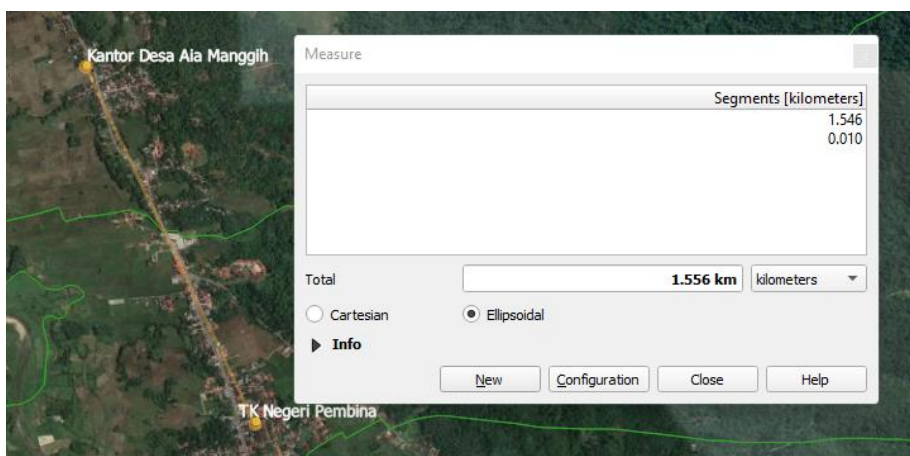


Figure 5. An illustration of measuring distance with the QGIS software's measure line feature

3.3. Score Calculation Results

Table 4. The result of calculating the score of indicators of the presence or access of urban facilities

Village	Existence or Access to Urban Facilities (Score)							Total Score
	A	B	C	D	E	F	G	
Aia Manggih	1	1	1	1	0	0	1	5
Aia Manggih Selatan	1	1	1	1	1	1	1	7
Aia Manggih Utara	0	1	0	1	0	0	1	2
Pauah	1	1	1	1	1	1	1	7
Durian Tinggi	1	1	1	1	1	1	1	7
Sundata Selatan	0	1	0	1	0	0	1	3
Jambak	1	1	1	1	1	1	1	7
Tanjung Beringin	1	1	1	1	0	0	1	5

A : Kindergarten School

B : Junior High School

C : Senior High School

D : Market

E : Shops

F : Hospital

G : Hotel/Billiards/Discotheque/Pub/Karaoke Venue/Salon

Table 4 above shows the results of calculating the score for the existence or access of urban facilities for eight villages in the Lubuk Sikaping District. Overlaying landmark points of urban facilities with digital maps of villages, as well as the method of calculating scores described in Table 2 earlier, form the basis for calculating scores in Table 4. Based on Table 4 above, it can be seen that four villages have the most significant potential to be classified into urban villages because they have a maximum score value, namely Aia Manggih Selatan, Pauah, Durian Tinggi, and Jambak. Furthermore, two villages have the most significant potential to be classified into rural villages because they have a minimum score: Sundata Selatan and Aia Manggih Utara. The final score will still be added with two other indicators, namely population density, agricultural families' percentage, and families using landlines and electricity.

4. CONCLUSION

This study calculated the classification score of urban and rural villages for indicators of the presence or access of urban facilities. BPS - Statistics Indonesia obtained data on urban facilities for classifying urban and rural villages through the Podes survey. The urban facilities that must be data on the Podes survey are quite a lot, so it takes a lot of time, cost, and energy. Therefore, this study tried to survey the use of landmark data on urban facilities on the google maps site to calculate the score of the existence or access of urban facilities.

Landmark data was obtained using web scraping methods through Botsol software. The data is then further processed using QGIS software, such as overlaying landmark points with a digital map of the village. Eight villages in Lubuk Sikaping Subdistrict were selected as case studies for this study and successfully calculated scores for indicators of the presence or access of urban facilities. Based on these scores, four large potential villages are classified into urban villages, and three likely villages are classified into rural villages. Future research is needed to develop a prototype of a smart system that can automatically extract data from the google maps site using the web scraping method and calculate the real-time score of the presence or accessibility of urban facilities.

REFERENCES

- [1] Badan Pusat Statistik, *Peraturan Kepala Badan Pusat Statistik Nomor 120 Tahun 2020 Tentang Klasifikasi Desa Perkotaan dan Perdesaan di Indonesia 2020 Buku 1 Sumatera*. Badan Pusat Statistik, 2020.
- [2] S. Pascasarjana, "Kajian penentuan klasifikasi desa di indonesia shafa rosea surbakti," 2015.
- [3] C. Muehlethaler dan R. Albert, "Collecting data on textiles from the internet using web crawling and web scraping tools," *Forensic Sci. Int.*, vol. 322, hal. 110753, 2021, doi: 10.1016/j.forsciint.2021.110753.
- [4] W. Srimulyani, N. Nurtia, M. Faris, N. F. Deli, dan S. Pramana, "Profil Tingkat Okupansi Hotel Di Ntb Selama Pandemi Covid-19 Dengan Menggunakan Big Data," *Semin. Nas. Off. Stat.*, vol. 2020, no. 1, hal. 273–280, 2021, doi: 10.34123/semnasoffstat.v2020i1.503.
- [5] S. Satriajati, S. B. Panuntun, dan S. Pramana, "Implementasi Web Scraping Dalam Pengumpulan Berita Kriminal Pada Masa Pandemi Covid-19," *Semin. Nas. Off. Stat.*, vol. 2020, no. 1, hal. 300–308, 2021, doi: 10.34123/semnasoffstat.v2020i1.578.
- [6] C. F. Annisa dan S. Pramana, "Kajian Pemanfaatan Data Google Maps dalam Official Statistics (Studi Kasus : Usaha Sektor Penyedia Makan Minum di Pulau Jawa dan Bali)," *Semin. Nas. Off. Stat.* 2020, vol. 2020, no. 1, hal. 328–337, 2020.
- [7] A. D. Churi, "Google Map Traffic Data Scraping and Mining," no. 04, hal. 891–895, 2021.
- [8] T. Yoshida, Y. Yamagata, dan D. Murakami, "Energy demand estimation using quasi-real-time people activity data," *Energy Procedia*, vol. 158, hal. 4172–4177, 2019, doi: 10.1016/j.egypro.2019.01.813.
- [9] S. Avetisyan, "Web Scraping and Geocoding : Some Stylized Facts About Regions of The Republic of Armenia Sergey Avetisyan To cite this version : HAL Id : hal-02015746 Web Scraping and Geocoding : Some Stylized Facts About Regions of The Republic of Armenia * List of Figures," 2019.

- [10] E. Uzun, "A Novel Web Scraping Approach Using the Additional Information Obtained from Web Pages," *IEEE Access*, vol. 8, hal. 61726–61740, 2020, doi: 10.1109/ACCESS.2020.2984503.
- [11] T. Karthikeyan, K. Sekaran, D. Ranjith, V. Vinoth kumar, dan J. M. Balajee, "Personalized content extraction and text classification using effective web scraping techniques," *Int. J. Web Portals*, vol. 11, no. 2, hal. 41–52, 2019, doi: 10.4018/IJWP.2019070103.
- [12] R. Diouf, E. N. Sarr, O. Sall, B. Birregah, M. Bousso, dan S. N. Mbaye, "Web Scraping: State-of-the-Art and Areas of Application," *Proc. - 2019 IEEE Int. Conf. Big Data, Big Data 2019*, hal. 6040–6042, 2019, doi: 10.1109/BigData47090.2019.9005594.
- [13] S. Sirisuriya, "A Comparative Study on Web Scraping," *8th Int. Res. Conf. KDU*, no. November, hal. 135–140, 2015.
- [14] R. Diouf, U. De Thies, dan U. Michel, "Web Scraping : State-of-the-Art and Areas of Application," hal. 6040–6042, 2019.