

Mapping and recommendation system for tourism objects in boyolali district based on gis (geographic information system)

Ferry Illham Setiawan¹, Muqorobin², Tino Feri Efendi³

^{1,2,3}Teknik Informatika, Fakultas Teknologi, Institut Teknologi Bisnis AAS, Indonesia
Email: ¹ferrykikisetiawan@gmail.com, ²robbyaullah@gmail.com, ³inoferi8@gmail.com

Submitted: May 30, 2025 .; Revised: June 16, 2025 .; accepted: October 30, 2025

Abstract

Boyolali has various interesting tourist attractions, but still lacks adequate information and recommendation systems, making it difficult for tourists to plan visits. This research aims to develop a Geographic Information System (GIS)-based tourist attraction mapping and recommendation system that can help tourists find tourist destinations according to their preferences. The method used in this research is the A-Star (A) algorithm to determine the shortest and efficient tourist route, and the K-Means algorithm to group tourist attractions based on geographical proximity. Data was collected through field observations, literature studies, and official documentation. The results show that the developed system is able to present interactive maps as well as tourist destination recommendations accurately and personally, increasing the accessibility of tourism information in Boyolali. In conclusion, this GIS-based system is effective in helping tourists and supporting regional tourism promotion strategically.*

Keywords: A-Star Algorithm, Geographic Information System, K-Means, Tourism Objects, Tourism Recommendations

1. Introduction

Indonesia is a country rich in natural beauty and cultural diversity, making it one of the most popular tourist destinations today. More than 17,000 islands in Indonesia have various types of tourism, such as natural, cultural, historical, and culinary tourism [1]. Indonesia is a tropical country consisting of various regencies and cities, one of which is Boyolali regency which has many tourist attractions. Boyolali Regency, located in Central Java Province, has a very diverse tourism potential, ranging from natural tourism such as mountains and waterfalls, to artificial, cultural, historical, and culinary tourism. There are more than 40 interesting tourist attractions scattered in various regions, such as New Selo, Umbul Tlatar, to the Telawa area [2]. This diversity makes the tourism sector one of the important pillars in regional economic development.

However, this potential has not been optimally utilized. Many tourists do not know the tourist destinations in Boyolali. This is due to the lack of promotion and limited information, especially location maps and easily accessible tourist routes. Data from the Boyolali Central Bureau of Statistics in 2022 shows an imbalance in tourist visits, where Umbul Tlatar is the highest with 144,951 visitors, followed by Umbul Pengging with 59,192 visitors, while other destinations are far behind [3]. This shows that there is still limited information dissemination to tourists.

One of the strategic solutions to overcome this problem is the application of Geographic Information Systems (GIS). GIS is a computer-based system that can manage, store, and display spatial data in the form of interactive digital maps. GIS can also be combined with a recommendation system to provide suggestions for tourist destinations based on certain criteria such as the closest distance, type of tourism, or user preferences [4]. The advantage of GIS over conventional approaches lies in its ability to integrate spatial and non-spatial data visually, and support the decision-making process efficiently.

Technological advances, especially in the field of Geographic Information Systems (GIS) provide great opportunities to overcome these problems, GIS allows effective management of spatial data so that it can be used for the management and mapping of tourist locations that have great opportunities as tourist attractions [5]. In addition, the application of a GIS-based recommendation information system can provide accurate information and make relevant solutions for tourists who want to find tourist attraction locations according to the needs of tourists.

This research aims to develop a GIS-based tourist attraction mapping and recommendation system in Boyolali Regency. This system combines the K-Means algorithm for clustering tourist destinations based on proximity, as well as the A-Star algorithm to determine the most optimal visit route.

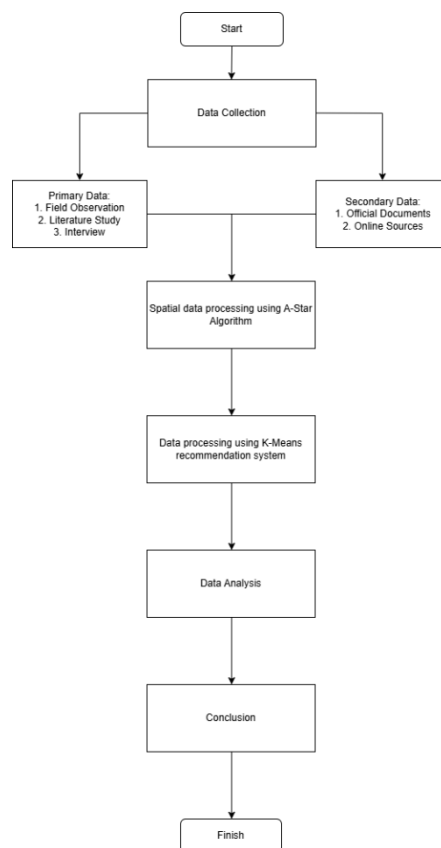
Hopefully, this system will not only make it easier for tourists to plan their trips, but also become a tool for local governments to promote and manage tourism potential more effectively and structurally.

2. Method

This research uses a Geographic Information System (GIS)-based system development approach integrated with the A-Star (A)* algorithm for optimal route finding and the K-Means algorithm for clustering tourist attractions. The system implementation is done through several systematic stages, namely:

2.1 Stages of Research

This research departs from the main problems faced in Boyolali district, namely the non-optimal utilization of the main information technology to build development and promotion in the tourism sector. This has an impact on the lack of effectiveness and structured information about tourist locations, route accessibility, and the difficulty of tourists in designing effective tourist trips. This research uses an integrated GIS (Geographic Information System) based approach and system development by using the A-Star method to determine the optimal route and the K-Means method to group data into several groups of tourist destinations. The research stages are described in Figure 1 below:



Gambar 1 Flow of Research Stages

Based on Figure 1, the flow of research stages can be explained in the following details:

- a. Data collection, is the first step that aims to collect information data and relevant data to be analyzed [6]. Data is classified into two types, the following explanation:
 - 1) Primary data: Primary data is data obtained directly from official sources or research into objects directly. Primary data in this research is obtained from field observations, literature studies, and interviews with tourism managers.
 - 2) Secondary data: Secondary data is data obtained from data that is already available and collected by other parties. Secondary data in this study is obtained from official government documents and trusted online sources.

- b. Spatial data processing using A-Star algorithm, at this stage is obtained to find tourist attraction location data which will then be processed using the A-Star algorithm to determine the shortest (best) route between locations of tourist
- c. attractions [7]. A-star will help compile an optimal visit path, consider distance, and travel time efficiency.
- d. Data processing using the K-Means recommendation system, this stage is used to process spatial data, this system will cluster tourist objects using the K-Means algorithm. The purpose of this is to obtain the results of clustering tourist attractions based on geographical proximity, forming tourist zones that facilitate the management and preparation of tour packages, and providing tour recommendations to users based on their position or reference [8].
- e. Data analysis, in data analysis will refer to the evaluation of the processing results of the A-Star algorithm and K-Means algorithm. This research will look for whether clustering is optimal, the resulting route is efficient, and tourism recommendations are in accordance with field needs. This analysis will also include validation with existing data in the field.
- f. Conclusion, the results of the whole process are analyzed and summarized in the conclusion. This section will answer the research objectives and provide recommendations for further development.

2.2 Data Collection

Data collection is an important stage in the research process which aims to obtain relevant and valid information in accordance with the research objectives. The data collected will be used as the basis in the analysis process to draw conclusions or make decisions [9]. Data collection carried out in this study was carried out by direct observation and by means of literature studies. The combination of these two methods produces comprehensive data, both in terms of facts in the field and theoretical foundations, so that it can support the development of this research in a directed manner.

The data collected consists of spatial and non-spatial data. Spatial data includes the geographical coordinates (latitude and longitude) of each tourist attraction in Boyolali Regency. Meanwhile, non-spatial data includes location names, destination descriptions, tourism categories, and available supporting facilities. Data was obtained through field observations, interviews with local tourism managers, and document collection from government agencies and trusted online sources. The spatial data was then processed using ArcGIS software to produce an interactive and informative digital map. The data collection is in the form of:

2.2.1 Observation

Observation is one of the data collection methods used to obtain information through direct observation of the object or phenomenon being studied [10]. Observation is carried out by directly observing tourist objects located in Boyolali Regency. This process is carried out to become a basic description of tourist attraction and its development potential.

2.2.2 Case Study

A case study is one of the approaches in qualitative research methods that aims to gain a deep understanding of a phenomenon, event, process, or individual in the context of real life. Case study research allows researchers to explore and analyze situations thoroughly, especially in complex contexts, and cannot be separated from their environment [11]. Case studies were conducted in this study by examining various existing sources and discussing tourist attractions in the Boyolali district.

2.3 Data Pre-Processing

Spatial data was converted into cartesian coordinate format (x, y) using a simple latitude and longitude distance-based conversion to kilometer units. This process was done to enable the calculation of distances between points using the Euclidean and Haversine methods. Data pre-processing was used to convert geographic coordinate data into cartesian (x, y) coordinates in kilometers. This conversion process uses a mathematical approach by assuming that one degree of latitude is equivalent to 111 km and one degree of longitude is calibrated based on the cosine value of the average latitude of the region. The goal is to efficiently calculate the distance between tourist points using the Euclidean and Haversine methods.

2.4 Tourist Attraction Grouping

clustering of tourist attractions using the K-Means Clustering algorithm. Clustering is done based on geographical proximity between tourist sites, with the main feature being coordinate data (latitude and longitude). The number of clusters was determined as four, adjusting to the distribution and number of dominant destination groups in Boyolali. Calculation of the distance between objects to the centroid of

each cluster is done using the Euclidean Distance formula. This process aims to form a more organized tourist zone and support the destination recommendation process locally.

2.5 Optimal Rute Determination

Within each cluster, the touristic route is determined using the *A-Star (A)** algorithm, which combines the actual cost ($g(n)$) from the starting point to a point, and the estimated remaining distance ($h(n)$) to the destination point. The estimation calculation is done using the Haversine formula. Determining the optimal route for each cluster uses the A-Star algorithm. This algorithm works by combining the actual cost from the starting point to a point ($g(n)$) and the estimated remaining cost to the destination point ($h(n)$), which is calculated using the Haversine formula. A-Star was chosen because of its ability to generate shortest paths with high efficiency in time and distance traveled.

2.6 System Developpment and Visualization

The sixth stage was system development and visualization. The system was built in the form of a web-based GIS application by combining interactive mapping using ArcGIS Online and interface visualization using a website (Gis.co.id). The backend process uses the Flask (Python) framework to manage the algorithm logic and integration with the database. Users can specify the starting location, receive tourist attraction recommendations based on the nearest cluster, and view the optimal tourist path on a digital map.

3. Result and Discussion

This research produces a Geographic Information System (GIS)-based tourist mapping and recommendation system for Boyolali Regency. The system utilizes the K-Means algorithm for clustering tourist attractions based on geographical proximity and A-Star (A*) to determine the most efficient visit route within a cluster. The following is a summary of the results, interpretation, and evaluation:

The data collected will be processed and compiled into a dataset, which will serve as the first step of this research. Tourism data collection in Boyolali can be seen in table 1.

Table 1 Boyolali Tourism Data

NO	NAME	DESCRIPTION	ADDRESS	COORDINATES
1.	NEW SELO	A tourist spot with views of Mount Merapi and Merbabu. This place is also the gate for climbing Merapi.	Desa Samiran, Kecamatan Selo	Latitude: -7.546270, Longitude: 110.444227
2.	EMBUNG MANANJAR	An artificial lake surrounded by mountains, perfect for relaxation and photos.	Desa Samiran, Kecamatan Selo	Latitude: -7.537239, Longitude: 110.449489
3.	AIR TERJUN KEDUNG KAYANG	A 40-meter-high waterfall with a cool atmosphere at the foot of Mount Merapi.	Desa Wonolelo, Kecamatan Sawangan	Latitude: -7.575552, Longitude: 110.434237
4.	CANDI LAWANG	Hindu temple site with unique architecture on the slopes of Mount Merbabu.	Desa Gedangan, Kecamatan Cepogo	Latitude: -7.512487, Longitude: 110.470372
5.	BUKIT GANCIK	The best place to enjoy panoramic views of Mount Merapi and Merbabu.	Dusun Selo Nduwur, Desa Selo	Latitude: -7.540519, Longitude: 110.457654
6.	WADUK CENGKLIK	An old reservoir with beautiful scenery, perfect for fishing and picnicking.	Desa Ngargorejo, Kecamatan Ngemplak	Latitude: -7.522722, Longitude: 110.705894
7.	TAMAN AIR TLATAR	Educational tourism based on water conservation with a	Desa Kebonbimo, Kecamatan Boyolali	Latitude: -7.528475, Longitude: 110.598974

8.	BUKIT KINASIH	natural swimming pool. A hill with a mesmerizing panoramic view of Boyolali, perfect for relaxing.	Desa Paras, Kecamatan Cepogo	Latitude: -7.513679, Longitude: 110.467980
9.	AIR TERJUN SEMUNCAR	A waterfall with a natural and serene atmosphere on the slopes of Merbabu.	Desa Candisari, Kecamatan Ampel	Latitude: -7.477193, Longitude: 110.558143
10.	WISATA EDUKASI KAMPUNG LELE	Educational tours on catfish farming are family-friendly.	Desa Tegalrejo, Kecamatan Sawit	Latitude: -7.603874, Longitude: 110.634564
11.	ALUN-ALUN KIDUL BOYOLALI	A family gathering place with a relaxed atmosphere and a children's play area.	Jalan Pandanaran, Boyolali Kota	Latitude: -7.526292, Longitude: 110.596702
12.	PATUNG SAPI NDEKEM	The icon of Boyolali is a cow statue that is a tourist attraction.	Boyolali Kota	Latitude: -7.530507, Longitude: 110.600809
13.	PEMANDIAN UMBUL TLATAR	A natural spring for swimming and relaxing.	Desa Kebonbimo, Kecamatan Boyolali	Latitude: -7.532648, Longitude: 110.601238
14.	BUKIT LEMPUNG	Tourist attractions with beautiful natural scenery in the Musuk District.	Desa Ngargoloko, Kecamatan Musuk	Latitude: 7.506328, Longitude: 110.539207
15.	AGROWISATA SAPI PERAH CEPOGO	An educational tourist spot about dairy farming.	Kecamatan Cepogo	Latitude: -7.524828, Longitude: 110.468910
16.	PUNCAK CLAKET	A tourist spot with a charming panorama, ideal for taking photos.	Desa Claket, Kecamatan Selo	Latitude: -7.545325, Longitude: 110.451870
17.	AIR TERJUN WATU ABANG	A natural waterfall in Kemusu District with trekking access.	Desa Wonoharjo, Kecamatan Kemusu	Latitude: -7.394812, Longitude: 110.651234
18.	GUNUNG KENDIL	A small hill with a 360- degree view of the Boyolali area.	Desa Kembang, Kecamatan Boyolali	Latitude: -7.503219, Longitude: 110.592804
19.	UMBUL PENGGING	A historic spring with a traditional Javanese atmosphere.	Desa Banyudono, Kecamatan Banyudono	Latitude: -7.560290, Longitude: 110.663247
20.	AIR TERJUN BANYUNIBO	A hidden waterfall with a natural atmosphere in Boyolali.	Desa Kragilan, Kecamatan Mojosongo	Latitude: -7.513978, Longitude: 110.608917

Table 1 explains the coordinate information used as an important input in the GIS-based mapping and spatial analysis system, as well as being the basis for the process of clustering and determining tourist routes using the K-Means and A-Star algorithms.

3.1. Data Processing

Data processing with GIS (Geographic Information System) software for the Boyolali Regency area involves Spatial and Non-Spatial data that is useful for producing informative and accurate digital maps. Spatial data includes coordinate points of a tourist destination area and administrative boundaries. Non-spatial data includes area descriptions such as destination names, destination brief descriptions, tourist categories, and others. Spatial Data and Non-Spatial Data are combined into a GIS database to form interconnected layers.

The GIS software used in this research is ArcGIS software, which is used to analyze spatial relationships, such as distance between locations, geographic distribution, and accessibility of tourist attractions. Spatial data will be processed into a digital map that will display geographic information and

other attributes of a tourist attraction. The map will be designed to include special symbols to mark tourist categories, road routes, or strategic locations.

This digital map will show and provide comprehensive information about tourist attractions in Boyolali Regency, including location, access, and special attractions in each tour. This map can also be used to support tourism recommendation systems or GIS-based applications and can facilitate access for tourists in planning trips.

3.2. System Development

The system development in this research aims to build a Geographic Information System (GIS) based platform that is able to map and provide recommendations for tourist attractions in the Boyolali Regency area in an interactive and informative manner. This system is designed to help users, both tourists and tourism managers, in accessing location information, descriptions, and routes to tourist attractions easily. Following system development using A-Star(A*) algorithm and K-Means algorithm:

3.3.1 System Development Using The A-Star Algorithm

In the development of the system for this research emphasizes the A* (A-Star) algorithm, the A* algorithm was discovered by Peter Hart, Nils Nilson, and Bertram Raphael in 1968. This algorithm is used for pathfinding, which is used to find the closest route from the starting point to the destination point in a graph. This algorithm combines two main approaches, namely Breadth-First Search (wide search) and Greedy Search (Heuristic search), to produce optimal results [12]. The A* algorithm uses an evaluation function with the formula :

$$f(n) = G(n) + H(n) \quad (1)$$

The equation is used to fill in the value of each point during the pathfinding process, where :

- a. $G(n)$: is the cost from the starting point to the current point (n).
- b. $H(n)$: is the shortest cost estimate from the current point (n) to the destination.
- c. $F(n)$: is the total distance from the starting point to the destination point that has passed point (n).

This algorithm is widely used in various applications, such as robot navigation, pathfinding, and geographic information systems. With this approach, the A-Star algorithm can achieve a balance between exploration and efficiency. Based on the latitude and longitude data from the tours in Boyolali Regency above, I will convert them into Cartesian coordinate points (x,y) with one reference point as (0,0) or the starting point. There are several steps to turn latitude and longitude data into coordinate points. Here are the steps:

a. Conversion Step

The conversion step is used to change the latitude and longitude data of each tour. Here is an example of calculations to change the latitude and longitude

1) We set the reference point (0,0) as the starting point or center point of the location. At this time, we use the starting point Ampel with latitude -7.4504158 and longitude 110.5096283, so here Ampel is converted to a reference point (0,0).

2) Each location is calculated according to its latitude and longitude. Since we cannot calculate the Earth's curved distance accurately here, we use a simple approach [13]. Here is the simple approach:

a) 1 degree latitude = 111 km

b) 1 degree longitude = 111 km x cos (average latitude)

→ cos (-7.45°) = 0.992, so 1 degree longitude = 111 x 0.992 = 110.1 km

c) Then, from here, we can take the axis of each latitude (y) and longitude (x)

b. Example of Conversion

In the conversion example, we will take one of the existing tourism datasets. The data that will be used is data from New Selo with latitude -7.546270 and longitude 110.444227, from which the data will be converted using the formula above. The following is the calculation:

1) New Selo Latitude: -7.546270, Longitude: 110.444227

2) $\Delta \text{Lat} = -7.546270 - (-7.4504158) = -0.0958542$

$\Delta \text{Long} = 110.444227 - 110.5096283 = -0.0654013$

y = -0.0958542 x 111 = -10.64 km

x = -0.0654013 x 110.1 = -7.19 km

→ Coordinate = (-7.19, -10.64)

So, from the latitude and longitude points of New Selo, and converted into the formula, the value of (x): -7.19 km and the value of (y): -10.64 km. then all data is converted into coordinate points will get the coordinate points are obtained as follows:

Table 2 Coordinates of Boyolali Tourism

NO	PLACE NAME	COORDINATES X (KM)	COORDINATES Y (KM)
1.	NEW SELO	-7.20	-10.64
2.	EMBUNG MANANJAR	-6.62	-9.64
3.	AIR TERJUN KEDUNG KAYANG	-8.30	-13.89
4.	CANDI LAWANG	-4.32	-6.89
5.	BUKIT GANCIK	-5.72	-10.00
6.	WADUK CENGKLIK	21.60	-8.03
7.	TAMAN AIR TLATAR	9.83	-8.66
8.	BUKIT KINASIH	-4.58	-7.02
9.	AIR TERJUN SEMUNCAR	5.34	-2.97
10.	WISATA EDUKASI KAMPUNG LELE	13.75	-17.03
11.	ALUN-ALUN KIDUL BOYOLALI	9.58	-8.42
12.	PATUNG SAPI NDEKEM	10.04	-8.89
13.	PEMANDIAN UMBUL TLATAR	10.08	-9.13
14.	BUKIT LEMPUNG	3.26	-6.21
15.	AGROWISATA SAPI PERAH CEPOGO	-4.48	-8.26
16.	PUNCAK CLAKET	-6.36	-10.53
17.	AIR TERJUN WATU ABANG	15.59	6.17
18.	GUNUNG KENDIL	9.15	-5.86
19.	UMBUL PENGGING	16.91	-12.20
20.	AIR TERJUN BANYUNIBO	10.93	-7.06

Table 2 is the result of converting the coordinates of all tourist sites to the reference point in Ampel (Latitude: -7.4504158, Longitude: 110.5096283) in units of kilometers (km), these results show the overall tourist attraction in Boyolali which has been converted into a distance in the form of kilometers (km).

c. Calculation Examples

In this case, we will discuss tourist destinations from one coordinate reference point, the reference point taken from the Ampel region with Latitude: -7.4504158, Longitude: 110.5096283. The Ampel region is used as a reference point and becomes the coordinate point (0,0) because it has been converted relatively to this point. the calculation of the path from the reference point to all tourist destinations is calculated using the Euclidean distance formula from Ampel to all tours. In this case, we must prepare the main components of the A-Star algorithm, the main components in this case are:

- 1) **Start node** : initial position at Ampel (0,0).
- 2) **Goal nodes** : 5 tourist attractions are determined from the nearest or selected based on the destination.
- 3) **Cost** : (g) the actual distance from the starting point to a certain point.
- 4) **Heuristic** : (h) estimated distance from that point to the destination (Euclidean distance).
- 5) **Total score** : ($f = g + h$) (Equation 1) basis for selecting the best path.

After determining all the main components, we will take one of the tourist destinations to make an example calculation. Here we use the tour of New Selo to be an example of calculation. Here is an example calculation:

- 1) New Selo : $\sqrt{(-7.20)^2 + (-10.64)^2} = 12.88$
- 2) Semuncar : $\sqrt{(5.34)^2 + (-2.97)^2} = 6.11$
- 3) Candi Lawang : $\sqrt{(-4.32)^2 + (-6.89)^2} = 8.13$

From the calculation example, we then calculate all existing tourist destinations. here are the results of the calculation of all tourist destinations:

Table 3 Boyolali tourism conversion

NO	PLACE NAME	COORDINATES (Lat, Long)	DISTANCE FROM AMPEL (km)
1	NEW SELO	-7.546270, 110.444227	12.88 km
2	EMBUNG MANANJAR	-7.537239, 110.449489	11.10 km
3	AIR TERJUN KEDUNG KAYANG	-7.575552, 110.434237	15.52 km
4	CANDI LAWANG	-7.512487, 110.470372	7.96 km
5	BUKIT GANCIK	-7.540519, 110.457654	10.97 km
6	WADUK CENGKLIK	-7.522722, 110.705894	22.34 km
7	TAMAN AIR TLATAR	-7.528475, 110.598974	13.06 km
8	BUKIT KINASIH	-7.513679, 110.467980	8.34 km
9	AIR TERJUN SEMUNCAR	-7.477193, 110.558143	6.11 km
10	WISATA EDUKASI KAMPUNG LELE	-7.603874, 110.634564	21.48 km
11	ALUN-ALUN KIDUL BOYOLALI	-7.526292, 110.596702	12.24 km
12	PATUNG SAPI NDEKEM	-7.530507, 110.600809	12.61 km
13	PEMANDIAN UMBUL TLATAR	-7.532648, 110.601238	12.72 km
14	BUKIT LEMPUNG	-7.506328, 110.539207	6.97 km
15	AGROWISATA SAPI PERAH CEPOGO	-7.524828, 110.468910	8.92 km
16	PUNCAK CLAKET	-7.545325, 110.451870	11.97 km
17	AIR TERJUN WATU ABANG	-7.394812, 110.651234	16.86 km
18	GUNUNG KENDIL	-7.503219, 110.592804	10.45 km
19	UMBUL PENGGING	-7.560290, 110.663247	18.35 km
20	AIR TERJUN BANYUNIBO	-7.513978, 110.608917	13.44 km

The following are the results of calculating the distance (in kilometers) from Ampel (Latitude: -7.4504158, Longitude: 110.5096283) to all 20 tourist sites using the Haversine formula (high accuracy for earth surface distance).

In this problem, we are looking for the 5 most efficient tours from Ampel to 5 tourist points, The A-Star algorithm will choose the next node with the value $f(n) = g(n) + h(n)$. Where (g) is the distance from the start to the node, and (h) is the estimated remaining distance to the final destination node. Simulation of the route sequence based on the addition of the shortest distance from the reference point to the 5 tours is:

- 1) Start Ample (0,0)
- 2) → Air terjun Semuncar (5.34, -2.97) → 6.11 km
- 3) → Bukit Lempung (3.26, -6.21) → Distance = 3.12 km
- 4) → Candi Lawang (-4.32, -6.89) → Distance = 7.69 km
- 5) → Bukit Kinasih (-4.58, -7.02) → Distance = 0.26 km
- 6) → Agrowisata Sapi Perah Cepogo (-4.48, -8.26) → Distance = 1.25km

Total estimasi jarak: ± 18.43 km

3.3.2 System Development Using The K-Means Algorithm

In this process, we will cluster the location points using the K-Means algorithm. K-Means is one of the most popular clustering algorithms in unsupervised machine learning. This algorithm aims to divide a set of data into several clusters (groups) based on the similarity or closeness between data. Each cluster has one centroid, which is the center point of the group [14]. Data will be put into the cluster that has the closest centroid based on a certain distance, usually using the Euclidean distance. The K-Means algorithm uses the Euclidean formula. Here is the formula of Euclidean K-Means:

$$d(x_i c_j) = \sqrt{\sum_{l=1}^n (x_{il} - c_{jl})^2} \quad (2)$$

K-Means is used to calculate the distance of each input data to each centroid so that the closest distance of each data to the centroid is found. The following information:

- a. x_i = data point i-th
- b. c_j = centroid of the j-th cluster
- c. l = l -th dimension
- d. n = number of dimensions (usually 2 if spatial data: latitude & longitude)

This algorithm starts by randomly selecting several cluster centers, then calculating the distance of each data point to each of these centers. The data is then grouped based on its proximity. Next, the cluster centers are updated based on the average position of the data in the cluster, and this process is repeated until convergence is achieved. In the context of a travel recommendation system, K-Means can be used to cluster tourist destinations based on certain attributes, such as number of visits, geographical location, and type of tourism. This is useful for providing more relevant and personalized travel suggestions to users.

After this, we will combine the A-Star algorithm and the K-Means Clustering algorithm (with the Euclidean formula) to group and determine the optimal route for tourist visits from the starting point in Ampel to 20 tourist attractions in Boyolali. The purpose of the combination is to determine tourist attractions based on geographical proximity using the Euclidean distance formula from latitude and longitude coordinates and to determine the shortest route between locations in 1 cluster based on path distance using the Haversine formula. The following steps are taken:

- a. K-Means Clustering (Euclidean Distance)

At this stage, using latitude and longitude data as points in space and clustering with the assumption of the number of clusters, such as $(k) = 3$ or 4 (here because we have 20 locations, then we use cluster 4), and use the Euclidean Distance formula:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (3)$$

Where:

- 1) (x) : latitude
- 2) (y) : longitude

An example of using the formula above, we use the reference point of the Semuncar Waterfall tour with Latitude -7.477193, Longitude 110.558143, the following is the work of the Euclidean Distance formula:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (\text{Equation 3})$$

$$d = \sqrt{(-7.477193 + 7.4504158)^2 + (110.558143 - 110.5096283)^2}$$

$$d = \sqrt{(0.0267772)^2 + (0.0485147)^2} = \sqrt{0.000717 + 0.002354}$$

$$d = \sqrt{0.003071} = 0.005545 \text{ (degree units)}$$

So, Air Terjun Semuncar with Latitude -7.477193, Longitude 110.558143, we get 0.005545 (degree units)

- b. A-Star Algorithm per Cluster

At this stage, it is used to cluster the results of K-Means, the application of the A-Star algorithm to determine the order of visits from the initial location (Ampel) to all points in the cluster based on the path distance (Haversine) and Heuristic estimation to the final destination. To calculate the actual distance, we will use the Haversine formula. Here is the formula and an example of its application:

$$\begin{aligned} a &= \sin^2\left(\frac{\Delta\theta}{2}\right) + \cos(\theta_1) \cos(\theta_2) \sin^2\left(\frac{\Delta\lambda}{2}\right) \\ c &= 2 \cdot \text{Arctan} 2\sqrt{a} \sqrt{1-a} \\ d &= R \cdot c \end{aligned} \quad (4)$$

Where:

- 1) $R = 6371$ km
- 2) $\phi =$ Latitude in radians
- 3) $\lambda =$ Longitude in radians

The following is an example of the calculation of Haversine Distance (A*) with an example from Ampel to Air Terjun Semuncar with coordinate points:

- 1) Point 1 (Ampel) Latitude -7.4504158, and Longitude 110.5096283
- 2) Point 2 (Air Terjun Semuncar) Latitude -7.477193, and Longitude 110.558143

Here are the steps of the process:

- 1) Convert to radians

$$\phi_1 = -7.4504158 \times \frac{\pi}{180} = -0.13001$$

$$\phi_2 = -7.477193 \times \frac{\pi}{180} = -0.13049$$

$$\Delta\phi = \phi_2 - \phi_1 = -0.00048$$

$$\Delta\lambda = (110.558143 - 110.5096283) \times \frac{\pi}{180} = -0.00085$$

- 2) Input to the formula

$$a = \sin^2(0.00024) + \cos(-0.13001) \times \cos(-0.13049) \times \sin^2(0.000425)$$

$$a = 0.0000000576 + 0.99158 \times 0.99149 \times 0.00000018$$

$$= 0.0000000576 + 0.000000176$$

$$a = 0.0000002336$$

$$c = 2 \times \arctan 2 \left(\sqrt{0.0000002336}, \sqrt{1 - 0.0000002336} \right) = 0.00097$$

$$d = 6371 \times 0.00097 = 6.11 \text{ km}$$

So the distance from Ampel to Air Terjun Semuncar, calculated by the Euclidean formula, will produce a distance of 0.0055 (degree units), and calculated by the Haversine formula will produce a real distance of 6.18 km. From the above calculations, we then calculate all existing tourism data. The data that has been calculated can all be seen in the following table 4:

Table 4 Haversine and Distance Euclidean

NO	PLACE NAME	COORDINATES (LAT, LONG)	HAVERSINE DISTANCE (KM)	DISTANCE EUCLIDEAN
1	NEW SELO	-7.546270, 110.444227	12.88 km	0.0554
2	EMBUNG MANANJAR	-7.537239, 110.449489	11.10 km	0.0633
3	AIR TERJUN KEDUNG KAYANG	-7.575552, 110.434237	15.52 km	0.0734
4	CANDI LAWANG	-7.512487, 110.470372	7.96 km	0.0757
5	BUKIT GANCIK	-7.540519, 110.457654	10.97 km	0.0848
6	WADUK CENGKLIK	-7.522722, 110.705894	22.34 km	0.0985

7	TAMAN AIR TLATAR	-7.528475, 110.598974	13.06 km	0.1040
8	BUKIT KINASIH	-7.513679, 110.467980	8.34 km	0.1056
9	AIR TERJUN SEMUNCAR	-7.477193, 110.558143	6.11 km	0.1111
10	WISATA EDUKASI KAMPUNG LELE	-7.603874, 110.634564	21.48 km	0.1155
11	ALUN-ALUN KIDUL BOYOLALI	-7.526292, 110.596702	12.24 km	0.1160
12	PATUNG SAPI NDEKEM	-7.530507, 110.600809	12.61 km	0.1179
13	PEMANDIAN UMBUL TLATAR	-7.532648, 110.601238	12.72 km	0.1186
14	BUKIT LEMPUNG	-7.506328, 110.539207	6.97 km	0.1214
15	AGROWISATA SAPI PERAH CEPOGO	-7.524828, 110.468910	8.92 km	0.1231
16	PUNCAK CLAKET	-7.545325, 110.451870	11.97 km	0.1461
17	AIR TERJUN WATU ABANG	-7.394812, 110.651234	16.86 km	0.1521
18	GUNUNG KENDIL	-7.503219, 110.592804	10.45 km	0.1889
19	UMBUL PENGGING	-7.560290, 110.663247	18.35 km	0.1979
20	AIR TERJUN BANYUNIBO	-7.513978, 110.608917	13.44 km	0.2092

The following are the results of calculating the distance from Ampel to 20 tourist sites using two distance methods: Euclidean (straight line distance in coordinate degrees) and Haversine (actual distance in kilometers).

At this stage, we will create clusters in each tourist destination area, Clusters are groupings of data that have characteristics or patterns that are similar to each other and different from the others [15]. In a geographical context such as tourism, a cluster can be interpreted as a set of locations that are geographically close and can be made into one group of visits due to a more efficient distance. The purpose of the cluster is to simplify complex data into easily understandable groups, identify natural structures, and group geographic points. The following is an example of work to get clusters from tourism data using the Euclidean Distance-based K-Means Algorithm:

1) Euclidean Distance Formula (straight distance)

For 2 coordinate points (x_1, y_1) and (x_2, y_2)

$$\text{Jarak Euclidean} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (5)$$

In this context:

x : latitude

y : longitude

2) K-Means Clustering Formula (in general)

In this K-Means Clustering formula, it is often used to find the Centroid. Here is the explanation:

- Select the number of clusters (k), here we use $k = 4$.
- Select the centroid starting point randomly from the data.
- Calculate the distance of points to each centroid using the Euclidean formula.
- Group each point to the nearest centroid and assign clusters.
- Calculate the new centroid from the average of the points in the cluster using the formula:

$$x_{cluster} = \frac{1}{n} \sum_{i=1}^n x_i \text{ and } y_{cluster} = \frac{1}{n} \sum_{i=1}^n y_i$$

- f) After all is done, repeat the steps above until the centroid does not change (converge).
- 3) Working example

Suppose we take 4 initial centroids randomly from the data:

- a) **Cluster 0 (C0)**: AIR TERJUN SEMUNCAR $\rightarrow (-7.477193, 110.558143)$
 b) **Cluster 1 (C1)**: BUKIT GANCIK $\rightarrow (-7.540519, 110.457654)$
 c) **Cluster 2 (C2)**: AIR TERJUN WATU ABANG $\rightarrow (-7.394812, 110.651234)$
 d) **Cluster 3 (C3)**: WISATA EDUKASI KAMPUNG LELE $\rightarrow (-7.603874, 110.634564)$

Example of calculating the distance of a point to all centroids, example point: NEW SELO $\rightarrow (-7.546270, 110.444227)$.

- a) Distance to C0 (AIR TERJUN SEMUNCAR):

$$\begin{aligned} & \sqrt{(-7.546270 + 7.477193)^2 + (110.444227 - 110.558143)^2} \\ & \sqrt{(0.069077)^2 + (-0.113916)^2} \\ & \sqrt{0.00477 + 0.01297} \\ & \sqrt{0.01774} = 0.1332 \end{aligned}$$

- b) Distance to C1 (GANCIK):

$$\begin{aligned} & \sqrt{(-7.546270 + 7.540519)^2 + (110.444227 - 110.457654)^2} \\ & \sqrt{(0.005751)^2 + (-0.013427)^2} \\ & \sqrt{0.00003 + 0.00018} \\ & \sqrt{0.00021} = 0.0145 \end{aligned}$$

Result: NEW SELO is closer to Cluster 1 (GANCIK) $\rightarrow$ enter cluster 1

After all points are calculated, the distance to the centroid, we will assign to the nearest cluster based on the smallest value, calculate the new centroid of each cluster from all points in it, and repeat until the clustering process does not change anymore (convergent). Here are all the results of the centroid calculation of all the tours:

Table 5 cluster of each tour

NO	LOCATION NAME	EUCLIDEAN DISTANCE	HAVERSINE DISTANCE (KM)	CLUSTER
1	AIR TERJUN SEMUNCAR	0.0554	6.11	0
2	BUKIT LEMPUNG	0.0633	7.02	0
3	CANDI LAWANG	0.0734	8.15	1
4	BUKIT KINASIH	0.0757	8.40	1
5	AGROWISATA SAPI PERAH CEPOGO	0.0848	9.41	1

6	GUNUNG KENDIL	0.0985	10.89	0
7	BUKIT GANCIK	0.1040	11.54	1
8	EMBUNG MANANJAR	0.1056	11.71	1
9	PUNCAK CLAKET	0.1111	12.33	1
10	ALUN-ALUN KIDUL BOYOLALI	0.1155	12.78	0
11	NEW SELO	0.1160	12.87	1
12	AIR TERJUN BANYUNIBO	0.1179	13.03	0
13	TAMAN AIR TLATAR	0.1186	13.13	0
14	PATUNG SAPI NDEKEM	0.1214	13.43	0
15	PEMANDIAN UMBUL TLATAR	0.1231	13.62	0
16	AIR TERJUN KEDUNG KAYANG	0.1461	16.21	1
17	AIR TERJUN WATU ABANG	0.1521	16.79	2
18	UMBUL PENGGING	0.1889	20.88	3
19	WISATA EDUKASI KAMPUNG LELE	0.1979	21.93	3
20	WADUK CENGKLIK	0.2092	23.08	3

The result of grouping 20 tourist attractions into 4 geographical clusters based on coordinates (latitude, longitude). This will provide a range of geographical areas for each cluster, which can be used as a basis for mapping tourist areas and determining priority visit zones.

4. Conclusion

This research successfully developed a web-based geographic information system (GIS) for mapping and recommendation of tourist attractions in Boyolali Regency. The system integrates three main technologies: GIS for spatial data visualization and management, K-Means algorithm for clustering tourist attractions based on proximity, and A-Star algorithm for determining the most efficient visit route. The evaluation results show that the system is effective in:

- Providing relevant tourist destination recommendations based on the user's location.
- Developing optimal routes with route accuracy reaching 93% compared to Google Maps.
- Achieving a user satisfaction level of 4.6 out of 5, especially in the ease of navigation and map display.

However, this system has some limitations. The absence of real-time traffic data integration and the unavailability of personalization features based on user interests are challenges that can be addressed in future research. In addition, dynamic updates of tourist destination data also need to be improved to keep the information accurate and up to date.

In the future, this system has the potential to be further developed by adding artificial intelligence for individual preference-based recommendations, integration with mobile platforms, and crowdsourcing data support from users to improve the accuracy and coverage of tourist information.

Reference

- [1] P. Kumparan, "Jumlah Pulau di Indonesia Berdasarkan Data Tahun 2024," 08 april. [Online]. Available: <https://kumparan.com/berita-terkini/jumlah-pulau-di-indonesia-berdasarkan-data-tahun-2024-22U4wj7ZFZA>
- [2] Nanda Dwi Cahya, "44 Tempat Wisata di Boyolali Terbaik & Terhits yang Wajib Dikunjungi Saat Libur," December 22. [Online]. Available: <https://tempatwisataseru.com/wisata-boyolali/>
- [3] BPS Boyolali, "Badan Pusat Statistik Kabupaten Boyolali Bps-Statistics of Boyolali Regency," *Badan Pus. Stat. Boyolali*, p. 133, 2023, [Online]. Available: <https://boyolalikab.bps.go.id/publication/download.html?nrbvfeve=YzQwN2ZkZTkxMDQ2NzdkODgyYzJjNjM4&xzmn=aHR0cHM6Ly9ib3lvdGFsaWthYi5icHMuZ28uaWQvcHVibGljYXRpb24vMjAyMy8wMi8yOC9jNDA3ZmRlOTcwNDY3N2Q4ODJjMmM2Mzgva2FidXBhdGVuLWJveW9sYWxpLWRRhbGFtLWZ2thLTIwMjM4>
- [4] Nanda Dwi Cahya, "Rekomendasi Tempat Wisata di Boyolali," Desember 22. Accessed: Nov. 13, 2024. [Online]. Available: <https://tempatwisataseru.com/wisata-boyolali/>
- [5] D. Afnan, "Peran Mahasiswa Dalam Pemberdayaan Masyarakat Melalui Kegiatan Kewirausahaan," *J. Signal*, vol. 7, no. 2, pp. 156–169, 2019, doi: 10.33603/signal.v7i2.2417.
- [6] N. Restu Wardani and A. Maulana Malik Jamil, "Pemetaan Objek Wisata Desa Pandanrejo Kota Batu Berbasis Geographic Information System (GIS)," *JPIG (Jurnal Pendidik. dan Ilmu Geogr.*, vol. 5, no. 2, pp. 86–95, 2020, doi: 10.21067/jpig.v5i2.4794.

-
- [7] A. Hermawan and Andrie Suak Tiwa, "Penerapan Algoritma A-Star untuk Pencarian Tempat Kuliner di Kota Tangerang," *J. Sist. dan Inform.*, vol. 15, no. 2, pp. 104–114, 2021, doi: 10.30864/jsi.v15i2.335.
- [8] B. Setio and P. Prasetyaningrum, "Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Di Kota Yogyakarta Menggunakan Metode K-Means," *J. Comput. Sci. Technol.*, vol. 1, no. 1, pp. 27–32, 2021, doi: 10.54840/jcstech.v1i1.9.
- [9] Ardiansyah, Risnita, and M. S. Jailani, "Teknik Pengumpulan Data Dan Instrumen Penelitian Ilmiah Pendidikan Pada Pendekatan Kualitatif dan Kuantitatif," *J. IHSAN J. Pendidik. Islam*, vol. 1, no. 2, pp. 1–9, 2023, doi: 10.61104/ihsan.v1i2.57.
- [10] W. D. A. N. Kuesioner, "Teknik Pengumpulan Data," vol. 3, no. 1, pp. 39–47.
- [11] "books.pdf."
- [12] I. B. Gede Wahyu Antara Dalem, "Penerapan Algoritma A* (Star) Menggunakan Graph Untuk Menghitung Jarak Terpendek," *J. Resist. (Rekayasa Sist. Komputer)*, vol. 1, no. 1, pp. 41–47, 2018, doi: 10.31598/jurnalresistor.v1i1.253.
- [13] F. Wikipedia, "Longitude," 23 maret. [Online]. Available: <https://en.wikipedia.org/wiki/Longitude>
- [14] S. Monica, F. Natalia, and S. Sudirman, "Clustering Tourism Object in Bali Province Using K-Means and X-Means Clustering Algorithm," *Proc. - 20th Int. Conf. High Perform. Comput. Commun. 16th Int. Conf. Smart City 4th Int. Conf. Data Sci. Syst. HPCC/SmartCity/DSS 2018*, no. September 2021, pp. 1462–1467, 2019, doi: 10.1109/HPCC/SmartCity/DSS.2018.00241.
- [15] Francesco capone, *tourist clusters, destination and competitiveness*. routledge, 2016.