KNN AND WEBGIS CLASSIFICATION TO RECOMMEND MOUNTAIN LOCATION ACCORDING TO HIKER ABILITIES

[1]SHAGI HISYAM AL FATHONY, [2]ANI DIJAH RAHAJOE, [3]RIFKI FAHRIAL ZAINAL, [4]MAS NURUL HAMIDAH, [5]ARIF ARIZAL

Department of Informatics Engineering, Bhayangkara University Surabaya

Jl. Ahmad Yani No.114, Ketintang, Kec Gayungan, Kota SBY, Jawa Timur 60231

e-mail: [1]shagi@gmail.com, [2]anidj@gmail.com, [3]rifki@ubhara.ac.id

ABSTRACT

The increasing number of climbers has an impact on the need for a system that can recommend mountains for climbing according to the ability of climbers. This study aims to create a system that can help climbers determine the mountain according to their abilities. Researchers use one of the methods in data mining, namely classification, using the K = Nearest Neighbor (K-NN) algorithm.
This research has produced a web-based system where this system can classify and provide recommendations according to the ability of climbers. This system is equipped with a hiking trail map which is expected to help make it easier for climbers to choose the mountain they will climb.

Keywords: *Mountain, Climbing, Route, KNN, Data Mining*

1. INTRODUCTION

Indonesia is an archipelagic country with beautiful mountain ranges. This row of mountains has stunning natural beauty, it is this natural beauty that attracts many tourists to visit it. This is what causes the emergence of many outdoor activities activists in Indonesia. This increase is also influenced by the number of influencers who also create content for outdoor activities, especially mountain climbing.

The increase in mountaineers is not accompanied by adequate information and physical preparation. Knowledge of the mountains to be visited, as well as physical abilities are quite important. Both of these are necessary in order to minimize accidents that may occur.

To make it easier to determine the mountain according to the physical abilities and experience of prospective climbers, we can use classification techniques in data mining. Data mining is a series of processes to explore added value from a data set in the form of knowledge that has not been known manually (Pramudiono, 2006). Classification is used to classify and determine which mountains are suitable for climbers' abilities.

By applying classification techniques in data mining, this study aims to provide mountain recommendations according to the abilities of prospective climbers using the K-Nearest Neighbor method. Based on these problems, the title "KNN Classification and Web GIS to determine recommendations for mountain climbing locations according to the ability of climbers.

2. WEBGIS

WebGIS is a GIS or digital mapping application that utilizes the internet network as a communication medium that functions to distribute, publish, integrate, communicate and provide information in the form of text, digital maps and perform analysis and query functions related to GIS through the internet network (Prahasta, 2010). 2007). webhGIS is designed to work with spatially referenced data or geographic coordinates. Here are some of the advantages of using WebGIS:

1. Users do not need special software, just use an internet browser such as Internet Explorer, Mozilla Fire Fox, Google Chrome and so on
2. Does not depend on the operating system so that it can be operated on all computers with various operating systems

3. K-NEAREST NEIGHBOR

The KNN algorithm is a classification algorithm that works by taking a number of K nearest data (neighbors) as a reference to determine the class of new data. This algorithm classifies data based on similarity or similarity or proximity to other data. In K-Nearest Neighbor, data points that are close together are called "neighbors" or "neighbors".

4. SYSTEM DEVELOPMENT

For now, mountain climbing information for climbers, especially novice climbers, is still limited. Many beginner climbers are often wrong in choosing a mountain to climb, so things often happen that are not desirable. This happened because of the lack of experience of the climbers who were not in accordance with the mountain terrain to be traversed. This problem is the background of the author to create a system that can recommend mountains that climbers climb.

The data used in this study came from the results of a survey using Google Forms of approximately 1000 data taken from mountaineer forums.

The system starts by logging in as admin. Admin can manage input maps, training data, and test data that will be entered into the database. After the training data is collected, the system will do pre-processing. Then the admin will process the test data input. The test data will be pre-processed first so that it can be tested.

After entering the test data, the next thing to do is to input the value of K as the value of the number of neighbors. Then the system will calculate the distance using the Euclidean distance formula. Then it will display the distance between the test data and all the training data in order from the smallest to the largest distance. Then the top data is taken according to the K value, and displays the results of mountain recommendations. Then you can see the recommended route from the mountain using maps.

The K value parameter can determine the result of the classification. In the classification process using K-NN the value of K determines the accuracy of the classification. For the use of more than one K parameter, voting by majority is used to determine the classification results.

The system testing will be carried out using a confusion matrix. Basically the confusion matrix provides information on the comparison of the classification results carried out by the system with the actual classification results. The confusion matrix is a matrix table that describes the performance of the classification model on a series of testing data whose actual value is known. The following table of confusion matrix and the Confucian Matrix formula to find the level of accuracy, precision, and recall.

Table 1. Confusion Matrix

Amout of Data	Actual Values	
Predicted Values	TP (True Positive)	FP (False Positive)
	FN (False Negative)	TN (True Negative)

Where as:

1. TP is True Positive, ie the number of positive data that is correctly predicted by the system.
2. TN is True Negative, ie the number of negative data that is correctly predicted by the system.
3. FN is False Negative, which is the amount of negative data but is predicted to be wrong by the system.
4. FP is False Positive, that is, the number of positive data but is predicted to be wrong by the system.

The Confusion Matrix formula can be seen in the equations of formulas 1, 2 and 3 as follows:

$$Accuration = \frac{TP+TN}{TP+TN+FP+FN}\ x\ 100\% \quad (1)$$

$$Precision = \frac{TP}{TP+FP}\ x\ 100\% \quad (2)$$

$$Recall = \frac{TP}{TP+FN}\ x\ 100\% \quad (3)$$

5.TEST RESULT

Testing the system carried out experiments as many as 50 data testing. The test is carried out by determining the value of the nearest neighbour, namely K = 7, K = 10, K = 15, K-20 against 50 testing data. After testing, the classification results will be obtained in the form of mountain recommendations according to the ability of climbers in accordance with previously determined variables. This test is done by comparing the results of the system with the results of manual calculations. The results of the comparison can be seen in table 2, and calculations will be carried out to determine the level of accuracy, precision, and recall of the results of these calculations.

Table 2 Test Result

Name	Age	Gender	PP	DK	DP	AI	PB	Real Data	System Data	Match
Alya Sapphire	24	Woman	0	45	0	7	Yes	Prau Mt.	Prau Mt.	Yes
Mellvia Isyani	19	Woman	1	45	2	5	No	Batur Mt.	Batur Mt.	Yes
Hesty	19	Woman	1	60	2	7	No	Bromo Mt.	Bromo Mt.	Yes
Dyah	22	Woman	5	60	2	4	No	Merbabu Mt.	Arjuno Mt.	No
Eny S	20	Woman	4	40	1	7	No	Merbabu Mt.	Merbabu Mt.	Yes
Rizka	20	Woman	2	30	1	8	No	Prau Mt.	Prau Mt.	Yes
Rani	19	Woman	4	80	3	8	No	Rinjani Mt.	Rinjani Mt.	Yes
Muhammad Nur Daim Masum	26	Man	12	30	2	8	No	Lawu Mt.	Lawu Mt.	Yes
Reyhand P	19	Man	2	20	2	5	No	Prau Mt.	Prau Mt.	Yes
Mahmud Rivan Hidayatullah	21	Man	3	120	4	4	No	Buthak Mt.	Buthak Mt.	Yes
Septianis Yuni Ayunda	19	Woman	0	45	0	5	No	Penanggungan Mt.	Batur Mt.	No
Diah Ayu Sari	18	Woman	0	50	0	4	No	Pundah Mt.	Batur Mt.	No
Hanief Aulia Rahman	19	Man	1	100	3	6	No	Cikuray Mt.	Cikuray Mt.	Yes
Rezki W	21	Man	5	20	1	8	No	Lawu Mt.	Lawu Mt.	Yes
Fajar Amri	20	Man	0	30	0	4	Yes	Prau Mt.	Prau Mt.	Yes
Yuli Pratiwi	22	Woman	2	15	0	6	No	Penanggungan Mt.	Penanggungan Mt.	Yes
Ikbal Ihza Mahendra	22	Man	0	15	0	3	No	Arjuno Mt,	Merapi Mt.	No
Ardian Masrur	21	Man	0	10	0	15	Yes	Andong Mt.	Andong Mt.	Yes
Ahmad Aris Abdillah	27	Man	15	60	3	10	No	Lawu Mt.	Lawu Mt.	Yes
Nuraini Meina	21	Woman	0	30	0	9	No	Telemoyo Mt.	Telemoyo Mt.	Yes
Eviona Vivan	18	Woman	2	30	3	2	No	Prau Mt.	Prau Mt.	Yes
Sollahur	20	Woman	2	45	2	2	Yes	Andong Mt.	Andong Mt.	Yes
Diana D	19	Woman	3	60	1	8	No	Bromo Mt.	Bromo Mt.	Yes
Feri Adrian	25	Man	35	100	2	9	No	Slamet Mt.	Slamet Mt.	Yes

Aela	24	Woman	2	120	1	6	No	Ijen Mt.	Ijen Mt.	Yes
Clara Desiree	21	Woman	3	35	2	5	No	Prau Mt.	Prau Mt.	Yes
Prayogo A	29	Man	2	30	1	6	No	Bromo Mt.	Bromo Mt.	Yes
Rey Dermawan	22	Man	10	20	3	6	No	Prau Mt.	Prau Mt.	Yes
Anggi	23	Man	4	50	3	6	Yes	Rinjani Mt.	Bromo Mt.	No
Ari	22	Woman	6	75	2	4	No	Slamet Mt.	Slamet Mt.	Yes
Fitri	22	Woman	1	45	1	15	No	Prau Mt.	Prau Mt.	Yes
Herlina Kusumaningtyas	20	Woman	1	90	2	5	No	Cikuray Mt.	Cikuray Mt.	Yes
Ando	27	Man	6	90	3	12	Yes	Buthak Mt.	Buthak Mt.	Yes
Imma S	27	Woman	1	15	1	8	No	Penanggungan Mt.	Penanggungan Mt.	Yes
Azizah Astri Wulandari	18	Woman	0	140	0	13	No	Prau Mt.	Lawu Mt.	No
Tri Umi Alwanisari	22	Woman	0	30	0	8	No	Kembang Mt.	Kembang Mt.	Yes
Onky F	25	Man	3	120	2	8	No	Dempo Mt.	Penanggungan Mt.	No
Prima Nurul Kamariyah	19	Woman	12	70	2	9	No	Welirang Mt.	Welirang Mt.	Yes
Ayu	21	Woman	0	120	0	8	No	Sumbing Mt.	Penanggungan Mt.	No
Bayu Maharani	20	Woman	0	30	0	9	No	Guntur Mt.	Papandayan Mt.	No
Salwa Audila	23	Woman	0	20	0	2	No	Andong Mt.	Andong Mt	Yes
Regina	22	Woman	0	45	0	5	No	Andong Mt.	Andong Mt.	Yes

From the comparison of the results obtained in table 2, calculations will be carried out to measure the performance of the system with a confusion matrix on the value of K.

Table 3. Matriks Confusion with K=15

	Prau Mt	Batur Mt.	Bromo Mt.	Merbabu Mt.	Lawu Mt	Andong Mt.	Rinjani Mt.	Buthak Mt.	Penanggungan Mt.	Pundak Mt.	Ijen Mt.	Cikuray Mt.	Arjuno Mt.	Sumbing Mt.	Dempo Mt.	Telemoyo Mt.	Slamet Mt.	Guntur Mt	Kembang Mt.	Welirang Mt.	Semeru Mt.	Merapi Mt.	Papandayan Mt8
Prau Mt	8				3																		
Batur Mt.		1																					
Bromo Mt.			5									1											
Merbabu Mt.				2									1										
Lawu Mt					3																		
Andong Mt.						4																	

Rinjani Mt.			1				1															
Buthak Mt.								2														
Penanggungan Mt.		1							3													
Pundak Mt.		1								0												
Ijen Mt.											1											
Cikuray Mt.												2										
Arjuno Mt.													1									
Sumbing Mt.									1					0								
Dempo Mt.									1						0							
Telemoyo Mt.																1						
Slamet Mt.									1								2					
Guntur Mt.																		0				1
Kembang Mt.																			1			
Welirang Mt.																				1		
Semeru Mt.																					1	
Merapi Mt.																						0
Papandayan Mt.																						0

.

Table 3 is the confusion matrix from the test using the value of K = 15. The figure shows the confusion matrix for the multi-class classification. There are no positive and negative in the matrix confusion for multi-class. From table 2 it can be calculated the level of accuracy by using the following formula equation.

$$Accuration = \frac{39}{50}\ x\ 100\% = 0{,}78\ x\ 100\% = 78\%$$

The level of precision and recall needs to be calculated from each class one by one which will be shown in table 4.

Table 4. Precision and Recall with K=15

Mountain	Precision	Recall
Prau	$\frac{8}{8+1} = 0{,}88$	$\frac{8}{8+0} = 1$
Batur	$\frac{1}{1+0} = 1$	$\frac{1}{1+2} = 0{,}33$
Bromo	$\frac{5}{5+1} = 0{,}83$	$\frac{5}{5+1} = 0{,}83$
Merbabu	$\frac{2}{2+1} = 0{,}66$	$\frac{2}{2+0} = 1$
Lawu	$\frac{3}{3+0} = 1$	$\frac{3}{3+1} = 0{,}75$

Andong	$\frac{4}{4+0}=1$	$\frac{4}{4+0}=1$
Rinjani	$\frac{1}{1+1}=0{,}5$	$\frac{1}{1+0}=1$
Buthak	$\frac{2}{2+0}=0{,}88$	$\frac{2}{2+0}=1$
Penanggunngan	$\frac{3}{3+1}=0{,}75$	$\frac{3}{3+3}=0{,}5$
Pundak	$\frac{0}{0+1}=0$	$\frac{0}{0+0}=0$
Ijen	$\frac{1}{1+0}=1$	$\frac{1}{1+0}=1$
Cikuray	$\frac{2}{2+0}=1$	$\frac{2}{2+1}=0{,}66$
Arjuno	$\frac{1}{1+1}=0{,}5$	$\frac{1}{1+1}=0{,}5$
Sumbing	$\frac{0}{0+1}=0$	$\frac{0}{0+0}=0$
Dempo	$\frac{0}{0+1}=0$	$\frac{0}{0+0}=0$
Telemoyo	$\frac{1}{1+0}=1$	$\frac{1}{1+0}=1$
Slamet	$\frac{2}{2+1}=0{,}6$	$\frac{2}{2+0}=1$
Guntur	$\frac{0}{0+1}=0$	$\frac{0}{0+0}=0$
Kembang	$\frac{1}{1+0}=1$	$\frac{1}{1+0}=1$
Welirang	$\frac{1}{1+0}=1$	$\frac{1}{1+0}=1$
Semeru	$\frac{1}{1+0}=1$	$\frac{1}{1+0}=1$
Merapi	$\frac{0}{0+0}=0$	$\frac{0}{0+1}=0$
Papandayan	$\frac{0}{0+0}=0$	$\frac{0}{0+1}=0$

Table 4 is a calculation table for recall and precision. After calculating per class, the next step is to calculate precision and recall the total for all existing classes, as follows:

$$Precision = \frac{Precision\ for\ each\ class}{Number\ of\ Class}\ x\ 100\%$$

$$= \frac{0{,}88+1+0{,}83+0{,}66+1+1+1+0{,}5+1+0{,}75+0+1+1+0{,}5+0+0+1+0{,}66+0+1+1+1+0+0}{23} x100\%$$

$$= \frac{10}{23}\ x\ 100\% = 43\%$$

$$Recall = \frac{Recall\ for\ each\ class}{Number\ of\ Class}\ x\ 100\%$$

$$= \frac{1+0{,}33+0{,}83+1+0{,}75+1+1+1+0{,}5+0+1+0{,}66+0{,}5+0+0+1+1+0+1+1+1+0+0}{23} x100\%$$

$$= \frac{11}{23} x100\% = 47\%$$

For the accuracy of the K = 15 value against 50 testing data, it is 78%, for the precision value is 43%, and for the recall value is 47%.

6.CONCLUSION

The conclusions that can be drawn from this research include the following:

1. The K-Nearest Neighbor (K-NN) method has succeeded in helping the classification process of mountain climbing recommendations.
2. The results of mountain recommendations are based on the classification of training data that has been collected through surveys
3. Based on the tests that have been carried out on 20 testing data that have been compared between systems with manual calculations, the comparison results show an accuracy rate of 78%, precision of 43%, and recall of 47%

REFERENCES

[1] Berry, Michael J.A. dan Gordon S. Linoff. 2004. Data Mining Techniques for Marketing, Sales, Customer Relationship Management. Second Edition. Wiley Publishing, Inc.
[2] Larose, Daniel T. 2005. Discovering Knowledge in Data: An Introduction to Data Mining. John Willey & Sons, Inc.
[3] Ponniah, P. 2001. Datawarehouse Fundamentals: A Comprehensive Guide for IT Professional. John Willey & Sons, Inc.
[4] Pramudiono, I. 2006. Apa itu Data Mining? Dalam http://datamining.japati.net/cgi-bin/indodm.cgi. Diakses tanggal 26 Februari 2021.
[5] Turban, E., dkk. 2005. Decision Support Systems and Intelligent Systems. Yogyakarta: Andi Offset.
[6] Kristanto, Andri. 2008. Perancangan Sistem Informasi dan Aplikasinya. Yogyakarta: Gava Media.
[7] Wahana Komputer. 2003. Tip dan Trik Pemrograman Delphi 7.0. Yogyakarta: Andi Offset.
[8] Bahri, Kusnassriyanto Saiful & Wawan Sjachriyanto. 2008. Teknik Pemrograman Delphi Edisi Revisi. Bandung: Informatika.
[9] Fayyad, Usama. 1996. Advances in Knowledge Discovery and Data Mining. MIT Press.