

Implementation of the K-Nearest Neighbor (K-NN) Algorithm in Classification of Angora and Country Cats

Wiwin Andriana^{*1}, Reza Wisnumurti², Yuni Lestari³, Bonifacius Vicky Indriyono⁴

Dian Nuswantoro University, Penanggungan No. 41 A Bandar Lor Kediri, (0354) 2895777

*E-mail : 612202200081@mhs.dinus.ac.id^{*1}, 612202200065@mhs.dinus.ac.id²,
612202200074@mhs.dinus.ac.id³, bonifacius.vicky.indriyono@dsn.dinus.ac.id⁴*

**Corresponding author*

Dibyo Adi Wibowo⁵, Erika Devi Udayanti⁶

Dian Nuswantoro University, Penanggungan No. 41 A Bandar Lor Kediri, (0354) 2895777

E-mail : dibyoadiwibowo@dsn.dinus.ac.id⁵, erikadevi@dsn.dinus.ac.id⁶

Abstract – There are so many types of mixed cats from various cat breeds, many people find it difficult to identify and classify them. Therefore, we need a method that can classify the type of cat breeds. In this study the authors used the K-Nearest Neighbor algorithm to make it easier to recognize and classify cat breeds based on certain characteristics. The author took samples of 2 types, the Anggora race and the Kampung race. The implementation stage is to determine the euclidean distance and sort it, then determine the K-value to find the nearest neighbor. In testing, the authors used 50 training data and 50 test data with 6 attributes used, namely body shape, nose width, nose height, food type, hair type and hair length. The results of the classification of cat breeds using the k-NN method obtained an accuracy rate of 94% and an error rate of 6%.

Keywords – Cat Breeds, Classification, Algorithm, K-Nearest Neighbor, Euclidean.

1. INTRODUCTION

Cats are pets that everyone can have. Sometimes not all cat owners know the breed of the cat. This is due to the large number of cat breeds coupled with the large number of mixed cats from crosses from different races [1]. Until now, the lineage cat was recorded as a strain pure (*pure breed*) only 1% and the rest are cats of mixed breeds like stray/village cats [2] . Each cat breed has its own characteristics, but because of the large number of cross-breeding between cat breeds, determining the breed of a cat is much more difficult. One of the determinations of the type of cat breed can be done by observing body shape, nose width, nose height, type of food, type of fur, and length of cat's fur.

Therefore, we need an appropriate method to determine the type of cat breed. One of the data mining algorithms is *K-Nearest Neighbor* (k-NN) which can be used to solve *classification* and *regression problems* [3]. The KNN algorithm is what the author uses to classify cat breeds. By using indicators of body shape, nose width, nose height, type of food, type of fur, and length of fur. It takes the calculation of the *Euclidean distance* and sorting the shortest distance from the existing data. k-NN performs a classification on a data based on the K value that has been set previously. The value of K in this algorithm must use an odd value if it

is used for the classification process, unlike if it is used to predict the value of K, it can be an odd or even value [4].

Research was developed from research that has been done by previous researchers. Among them is the implementation of the *K-Nearest Neighbor* (k-NN) algorithm in determining the type of cat [5], in this study the attributes used are still few so that the accuracy is also smaller, while in this study the author used 6 attributes and the cat race studied was the Angora race. and village race. Then there are also other studies that apply the k-NN algorithm, namely the analysis of the *K-Nearest Neighbor* method in the classification of iris data [6], but the object under study is different from the object used by the author.

K-Nearest Neighbor algorithm is not only used for classification but is also widely used to solve prediction problems. One of them is the application of the k-NN method that has been carried out in research including predictions of the results of sawing sengon wood[7], predictions of rice yields in the Special Region of Yogyakarta Province using the k-NN method[8], predictions of on-time graduation of students[9], prediction of goods leaving the warehouse using the *time series method* k-NN[10], and prediction of the amount of *coconut oil production* using the k-NN and *Backward elimination methods* [11].

2. RESEARCH METHODS

2.1. K-Nearest Neighbor (k-NN) Algorithm

The *k-Nearest Neighbor* (k-NN) algorithm is a *supervised algorithm* that has *query results* new *instance* obtained by classifying the majority value of the category. The k-NN algorithm has the aim of classifying new objects based on data attributes and *training samples* [12]. Implementation of the k-NN algorithm by collecting datasets as *training data*, determining the value of k, entering new data as *testing data*, looking for the Euclidean distance, sorting the nearest neighbors according to the value of k, and classification is determined from the value that appears the most. The value of k is the number of nearest neighbors to determine the classification results. The distance value from the classroom is calculated from the dataset *marked by x^i training* and x^j testing of the new data values entered [3].

With the k-NN method, the classification of cat types can be done according to the characteristics of the cat. In the research conducted, the objects observed were the Angora cat and the village cat. In other words, this study was conducted to classify cats, whether cats with existing characteristics are included in the category of Angora cats or native cats.

The distance used is the calculation of the *Euclidean distance* namely the search method between two variable points, the closer and similar the smaller the distance between the two points [13]. *Euclidean Distance* is said to be good if the new data has a minimum distance and high similarity [14]. Calculations for measuring the distance between two points in *Euclidean space* study the relationship between angles and distances. This *Euclidean* distance calculation has the following formula:

$$d_i = \sqrt{\sum_{1=i}^n (x_{2i} - x_{1i})^2}$$

Description :

x_1 = sample test data

x_2 = test data
 i = data variable
 d = *dissimilarity* /distance
 n = dimension data

2.2. Research methods

In this case study of the k-NN method, the calculation process used is the Euclidean distance calculation process. The steps to calculate the K-Nearest Neighbor method with *Euclidean* distance include:

- Determining the K parameter : At this stage the determination of the K value has no certainty so that the value of K is relative. However, if there is a condition where the number of votes is balanced, then the value of K is reduced by 1 for each finding a balanced voting result.
- Calculating the *Euclidean distance* between the data to be evaluated with all training: At this stage, calculating the distance on k-NN is done by calculating each case data with the *Euclidean distance*.
- Sort the distances formed : After calculating each case with the *Euclidean distance* , the next step is to sort the data from the known quantities starting from the closest to the furthest.
- Determine the shortest distance to the K sequence: If the K value has been determined and the distance for each case has been calculated, then we take some cases according to the K value, starting from the smallest sequence.
- Pairing the appropriate classes : After the shortest distance to the sequence K has been determined, the classes are paired, which then looks for the number of classes from their nearest neighbours.
- Look for the number of classes from the nearest neighbor and define the class as the data class to be evaluated : to determine the class to be evaluated, the classes taken according to the k value are searched for in number from the nearest neighbors.

The research method used can be described as in Figure 1

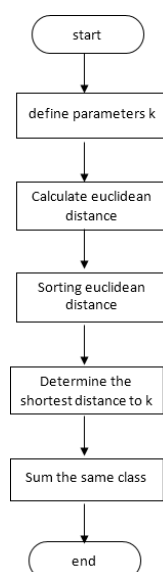


Figure 1. Research Methods

2.3. Accuracy Testing

Testing is done to know the accuracy of the method used. Accuracy is the degree of closeness of the predicted value to the actual calculated value [15]. The accuracy value is obtained by comparing the amount of data that is correct or wrong against the total amount of data (*testing*). The following is the accuracy calculation formula.

$$\text{Accuracy rate} = \frac{\sum \text{true value}}{\sum \text{overall data}} \times 100\%$$

Description :

\sum True value = The amount of data tested with correct results

\sum Overall data = Total data tested

3. RESULTS AND DISCUSSION

3.1. Processing of Training Data and Test Data

In this study, the authors collected 100 data consisting of 50 training data and 50 test data. By using 6 attributes, namely body shape, hair type, hair length, nose height, nose width, and type of food. Data from each attribute will be grouped as follows:

Table 1. Grouping of Body Shape Data Input

Class	Label
plump	1
slim	2

At table 1. Body shape is classified into 2, namely plump and slender. For the next plump classification will be labeled 1 and label 2 for slender classification.

Table 2. Grouping of Feather Type Input Data

Class	Label
Short straight	1
straight length	2
Solid length	3

At Table 2. The types of fur will be grouped into 3, namely short straight for label 1, long straight for label 2, long solid for label 3.

Table 3. Grouping of Feather Length Data Input

Class	Value limit	Label
Short	0-10mm	1
Currently	11-20mm	2
Long	21-30mm	3

At table 3. Feather lenght will be grouped into 3. Short feather with a value limit of 0-10 mm is labeled 1, medium feather with a value limit of 11-20mm is labeled 2, long feather with a value limit of 21-30mm.

Table 4. Grouping of Nose Height Data Input

Class	Value limit	Label
Low	1-3mm	1
Currently	4-6mm	2
Tall	7-9mm	3

At table 4 classifies the size of the nose height. Low class with a value limit of 1-3mm is included in label 1, Medium class with a value limit of 4-6mm is included in label 2, High class with a value limit of 7-9mm is included in label 3.

Table 5. Grouping of Nose Width Data Input

Class	Value limit	Label
Narrow	1-3mm	1
Currently	4-6mm	2
Wide	7-9mm	3

At table 5 classifies the width of the nose. Narrow noses with a value limit of 1-3mm fall under label 1, medium noses with a value limit of 4-6mm fall under label 2, wide noses with a value limit of 7-9mm fall under label 3.

Table 6. Grouping of Food Type Data Input

Class	Label
Dry Food	1
Wet Food	2
Fish	3
Chicken	4

In table 6 classifies the types of food. dry food is included in label 1, wet food is included in label 2, fish food is included in label 3, chicken food is included in label 4

The following is a training data table from the results of data collection.

Table 7 . Original Train Data

Case	Form Body	Feather Length	Nose Width	Nose Height	Food Type	Feather Type	Type of Cat
1	slim	16	6	6	Fish	straight short	Village
2	slim	18	7	7	fish	straight short	Village
3	plump	28	9	9	dry food	straight length	Angora
4	slim	5	3	3	wet food	straight short	Village
5	plump	21	6	6	chicken	straight length	Angora
6	plump	21	6	6	chicken	solid length	Angora
7	plump	25	7	6	chicken	straight length	Angora
8	slim	17	7	7	fish	straight short	Village
9	plump	26	8	8	dry food	solid length	Angora
10	slim	5	3	3	wet food	straight short	Village
11	slim	17	7	7	fish	straight short	Village
12	plump	26	8	8	dry food	straight length	Angora
13	slim	4	3	3	wet food	straight short	Village
14	plump	24	7	6	chicken	solid length	Angora
15	slim	16	6	6	fish	straight short	Village
16	slim	19	7	7	fish	straight short	Village
17	plump	27	9	8	dry food	solid length	Angora
18	plump	26	8	8	dry food	straight short	Angora
19	plump	26	8	9	dry food	straight length	Angora
20	plump	28	9	9	dry food	solid length	Angora
21	slim	5	3	3	wet food	straight short	Village
22	slim	16	6	6	fish	straight short	Village
23	plump	27	9	8	dry food	straight length	Angora
24	slim	5	3	3	wet food	straight short	Village
25	plump	23	7	7	chicken	straight length	Angora
26	slim	18	7	7	fish	straight short	Village
27	plump	29	9	9	dry food	solid length	Angora
28	plump	10	3	3	wet food	solid length	Angora
29	slim	18	7	7	fish	straight short	Village
30	plump	27	9	9	dry food	solid length	Angora
31	plump	8	3	2	wet food	solid length	Angora

32	plump	21	6	6	chicken	solid length	Angora
33	plump	22	6	6	chicken	solid length	Angora
34	plump	30	9	9	dry food	straight length	Angora
35	plump	29	9	9	dry food	straight length	Angora
36	slim	17	7	7	fish	straight short	Village
37	slim	17	7	7	chicken	straight short	Village
38	slim	19	7	7	fish	straight length	Village
39	plump	30	9	9	wet food	straight length	Angora
40	plump	28	9	9	dry food	solid length	Angora
41	plump	7	3	3	wet food	straight length	Angora
42	plump	22	6	6	chicken	solid length	Angora
43	plump	27	9	9	dry food	straight length	Angora
44	plump	16	6	6	fish	straight short	Angora
45	slim	20	7	7	fish	straight short	Village
46	slim	17	7	7	fish	straight short	Village
47	plump	30	9	9	dry food	solid length	Angora
48	plump	27	9	9	dry food	solid length	Angora
49	plump	26	8	8	dry food	solid length	Angora
50	slim	18	7	7	fish	straight short	Village

Table 8 . Original Test Data

Case	Form Body	Feather Length	Nose Width	Nose Height	Food Type	Feather Type
1	slim	27	9	9	dry food	straight length
2	plump	22	8	7	dry food	straight length
3	slim	17	7	7	fish	straight short
4	slim	9	3	2	wet food	solid length
5	slim	11	5	4	chicken	straight length
6	slim	16	9	9	fish	straight short
7	plump	30	7	7	dry food	straight length
8	plump	28	7	8	dry food	solid length
9	plump	7	3	2	wet food	solid length
10	slim	15	4	4	chicken	solid length
11	slim	17	8	8	fish	straight short
12	plump	29	8	7	dry food	solid length
13	slim	12	6	6	chicken	solid length
14	slim	12	6	5	chicken	straight length
15	plump	21	8	8	dry food	straight length
16	plump	21	9	9	dry food	solid length
17	plump	25	9	8	dry food	solid length
18	slim	9	3	3	wet food	straight length
19	plump	27	7	7	dry food	solid length
20	slim	20	8	8	fish	straight short
21	plump	25	9	8	dry food	straight length
22	slim	13	5	4	chicken	straight length
23	plump	8	3	2	wet food	solid length
24	slim	11	6	5	chicken	straight length
25	plump	26	9	8	dry food	straight length
26	plump	27	8	7	dry food	solid length
27	plump	30	8	8	dry food	solid length
28	slim	18	7	7	fish	straight short
29	slim	20	7	7	fish	straight short
30	slim	16	9	9	chicken	straight short
31	plump	29	9	8	dry food	solid length
32	slim	15	4	4	chicken	solid length
33	plump	9	3	3	wet food	solid length
34	slim	19	8	8	fish	straight short
35	plump	24	7	7	dry food	straight length
36	plump	23	7	7	dry food	solid length
37	plump	21	9	8	dry food	solid length
38	slim	20	8	8	fish	straight short
39	plump	22	8	7	dry food	solid length
40	plump	9	3	3	wet food	straight short

41	slim	20	7	7	fish	straight short
42	slim	19	9	9	fish	straight short
43	plump	25	9	8	dry food	straight length
44	slim	10	3	2	wet food	straight length
45	plump	25	7	6	dry food	straight length
46	plump	25	8	7	dry food	solid length
47	plump	26	9	7	dry food	solid length
48	slim	16	8	8	fish	straight short
49	slim	18	8	8	fish	straight short
50	slim	16	8	8	fish	straight short

If we look at the original training data and original test data above, each attribute has various data values, so simplification and grouping of the values of each attribute are needed. Following are the training data and test data after simplifying and grouping:

Table 9. Training Data After Grouping

Case	Body	Feather Length	Nose Width	Nose Height	Food Type	Feather Type	Type of Cat
1	2	2	2	2	3	1	2
2	2	2	3	3	3	1	2
3	1	3	3	3	1	2	1
4	2	1	1	1	2	1	2
5	1	3	2	2	4	2	1
6	1	3	2	2	4	3	1
7	1	3	3	2	4	2	1
8	2	2	3	3	3	1	2
9	1	3	3	3	1	3	1
10	2	1	1	1	2	1	2
11	2	2	3	3	3	1	2
12	1	3	3	3	1	2	1
13	2	1	1	1	2	1	2
14	1	3	3	2	4	3	1
15	2	2	2	2	3	1	2
16	2	2	3	3	3	1	2
17	1	3	3	3	1	3	1
18	1	3	3	3	1	1	1
19	1	3	3	3	1	2	1
20	1	3	3	3	1	3	1
21	2	1	1	1	2	1	2
22	2	2	2	2	3	1	2
23	1	3	3	3	1	2	1
24	2	1	1	1	2	1	2
25	1	3	3	3	4	2	1
26	2	2	3	3	3	1	2
27	1	3	3	3	1	3	1
28	1	1	1	1	2	3	1
29	2	2	3	3	3	1	2
30	1	3	3	3	1	3	1
31	1	1	1	1	2	3	1
32	1	3	2	2	4	3	1
33	1	3	2	2	4	3	1
34	1	3	3	3	1	2	1
35	1	3	3	3	1	2	1
36	2	2	3	3	3	1	2
37	2	2	3	3	4	1	2
38	2	2	3	3	3	2	2
39	1	3	3	3	2	2	1
40	1	3	3	3	1	3	1
41	1	1	1	1	2	2	1
42	1	3	2	2	4	3	1
43	1	3	3	3	1	2	1
44	1	2	2	2	3	1	1
45	2	2	3	3	3	1	2

46	2	2	3	3	3	1	2
47	1	3	3	3	1	3	1
48	1	3	3	3	1	3	1
49	1	3	3	3	1	3	1
50	2	2	3	3	3	1	2

Table 10. Test Data After Grouping

Case	Body	Feather Length	Nose Width	Nose Height	Food Type	Feather Type
1	2	3	3	3	1	2
2	1	3	3	3	1	2
3	2	2	3	3	3	1
4	2	1	1	1	2	3
5	2	2	2	2	4	2
6	2	2	3	3	3	1
7	1	3	3	3	1	2
8	1	3	3	3	1	3
9	1	1	1	1	2	3
10	2	2	2	2	4	3
11	2	2	3	3	3	1
12	1	3	3	3	1	3
13	2	2	2	2	4	3
14	2	2	2	2	4	2
15	1	3	3	3	1	2
16	1	3	3	3	1	3
17	1	3	3	3	1	3
18	2	1	1	1	2	2
19	1	3	3	3	1	3
20	2	2	3	3	3	1
21	1	3	3	3	1	2
22	2	2	2	2	4	2
23	1	1	1	1	2	3
24	2	2	2	2	4	2
25	1	3	3	3	1	2
26	1	3	3	3	1	3
27	1	3	3	3	1	3
28	2	2	3	3	3	1
29	2	2	3	3	3	1
30	2	2	3	3	4	1
31	1	3	3	3	1	3
32	2	2	2	2	4	3
33	1	1	1	1	2	3
34	2	2	3	3	3	1
35	1	3	3	3	1	2
36	1	3	3	3	1	3
37	1	3	3	3	1	3
38	2	2	3	3	3	1
39	1	3	3	3	1	3
40	1	1	1	1	2	1
41	2	2	3	3	3	1
42	2	2	3	3	3	1
43	1	3	3	3	1	2
44	2	1	1	1	2	2
45	1	3	3	2	1	2
46	1	3	3	3	1	3
47	1	3	3	3	1	3
48	2	2	3	3	3	1
49	2	2	3	3	3	1
50	2	2	3	3	3	1

3.2. Application of the K-Nearest Neighbor Algorithm

From the training data and test data, the type of cat will be determined using the K-Nearest Neighbor method with the following steps:

- From the existing training data and test data, the *Euclidean distance will be determined* first. It is known that the number of attributes in this study is 6 so that the *Euclidean formula* is as follows;

$$d(a,b,c,d,e,f) = \sqrt{(a_{1i}-a_{2i})^2 + (b_{1i}-b_{2i})^2 + (c_{1i}-c_{2i})^2 + (d_{1i}-d_{2i})^2 + (e_{1i}-e_{2i})^2 + (f_{1i}-f_{2i})^2}$$

Information :

d = euclidean distance

a_{1i} = training data from the i -th attribute

a_{2i} = test data from the i -th attribute

i = number of variables

With 50 training data and 50 test data and 6 attributes, namely body shape, hair type, hair length, nose height, nose width, and type of food. The following is the calculation of the *Euclidean distance* with the first 50 training data and test data. 1st test data with body shape 2, hair length 3, nose width 3, nose height 3, type of food 1 and type of fur2.

1st data euclidean distance

$$d1 = \sqrt{(2-2)^2 + (2-3)^2 + (2-3)^2 + (2-3)^2 + (3-1)^2 + (1-2)^2} = 2.82843$$

Euclidean distance of 2nd data

$$d2 = \sqrt{(2-2)^2 + (2-3)^2 + (3-3)^2 + (3-3)^2 + (3-1)^2 + (1-2)^2} = 2,44949$$

Euclidean distance of 3rd data

$$d3 = \sqrt{(1-2)^2 + (3-3)^2 + (3-3)^2 + (3-3)^2 + (1-1)^2 + (2-2)^2} = 1$$

- And so on the calculation is done up to 50 training data. The following is the *Euclidean distance* of all training data with the first training data.

Table 11. Euclidean distance results

Case	Body	Feather Length	Nose Width	Nose Height	Food Type	Feather Type	Type of Cat	Euclidean distance
1	2	2	2	2	3	1	2	2.82843
2	2	2	3	3	3	1	2	2,44949
3	1	3	3	3	1	2	1	1
4	2	1	1	1	2	1	2	3.74166
5	1	3	2	2	4	2	1	3,46410
6	1	3	2	2	4	3	1	3.60555
7	1	3	3	2	4	2	1	3.31662
8	2	2	3	3	3	1	2	2,44949
9	1	3	3	3	1	3	1	1.41421
10	2	1	1	1	2	1	2	3.74166
11	2	2	3	3	3	1	2	2,44949
12	1	3	3	3	1	2	1	1
13	2	1	1	1	2	1	2	3.74166
14	1	3	3	2	4	3	1	3.46410
15	2	2	2	2	3	1	2	2.82843
16	2	2	3	3	3	1	2	2.44949
17	1	3	3	3	1	3	1	1.41421
18	1	3	3	3	1	1	1	1.41421
19	1	3	3	3	1	2	1	1

20	1	3	3	3	1	3	1	1.41421
21	2	1	1	1	2	1	2	3.74166
22	2	2	2	2	3	1	2	2.82843
23	1	3	3	3	1	2	1	1
24	2	1	1	1	2	1	2	3.74166
25	1	3	3	3	4	2	1	3.16228
26	2	2	3	3	3	1	2	2.44949
27	1	3	3	3	1	3	1	1.41421
28	1	1	1	1	2	3	1	3.87298
29	2	2	3	3	3	1	2	2.44949
30	1	3	3	3	1	3	1	1.41421
31	1	1	1	1	2	3	1	3.87298
32	1	3	2	2	4	3	1	3.60555
33	1	3	2	2	4	3	1	3.60555
34	1	3	3	3	1	2	1	1
35	1	3	3	3	1	2	1	1
36	2	2	3	3	3	1	2	2.44949
37	2	2	3	3	4	1	2	3.31662
38	2	2	3	3	3	2	2	2.23607
39	1	3	3	3	2	2	1	1.41421
40	1	3	3	3	1	3	1	1.41421
41	1	1	1	1	2	2	1	3.74166
42	1	3	2	2	4	3	1	3.60555
43	1	3	3	3	1	2	1	1
44	1	2	2	2	3	1	1	3
45	2	2	3	3	3	1	2	2.44949
46	2	2	3	3	3	1	2	2.44949
47	1	3	3	3	1	3	1	1.41421
48	1	3	3	3	1	3	1	1.41421
49	1	3	3	3	1	3	1	1.41421
50	2	2	3	3	3	1	2	2.44949

- b. Smallest to largest *Euclidean* distance. The following table shows the results of the *Euclidean distance* set el ah sorted.

Table 12. Data After Euclidean Distance Sorted

Case	Body	Feather Length	Nose Width	Nose Height	Food Type	Feather Type	Type of Cat	Euclidean distance 1
3	1	3	3	3	1	2	1	1
12	1	3	3	3	1	2	1	1
19	1	3	3	3	1	2	1	1
23	1	3	3	3	1	2	1	1
34	1	3	3	3	1	2	1	1
35	1	3	3	3	1	2	1	1
43	1	3	3	3	1	2	1	1
9	1	3	3	3	1	3	1	1.41421
17	1	3	3	3	1	3	1	1.41421
18	1	3	3	3	1	1	1	1.41421
20	1	3	3	3	1	3	1	1.41421
27	1	3	3	3	1	3	1	1.41421
30	1	3	3	3	1	3	1	1.41421
39	1	3	3	3	2	2	1	1.41421
40	1	3	3	3	1	3	1	1.41421
47	1	3	3	3	1	3	1	1.41421
48	1	3	3	3	1	3	1	1.41421
49	1	3	3	3	1	3	1	1.41421
38	2	2	3	3	3	2	2	2.23607
2	2	2	3	3	3	1	2	2.44949
8	2	2	3	3	3	1	2	2.44949
11	2	2	3	3	3	1	2	2.44949
16	2	2	3	3	3	1	2	2.44949
26	2	2	3	3	3	1	2	2.44949

29	2	2	3	3	3	1	2	2,44949
36	2	2	3	3	3	1	2	2,44949
45	2	2	3	3	3	1	2	2,44949
46	2	2	3	3	3	1	2	2,44949
50	2	2	3	3	3	1	2	2,44949
44	1	2	2	2	3	1	1	2.82843
1	2	2	2	2	3	1	2	2.82843
15	2	2	2	2	3	1	2	2.82843
22	2	2	2	2	3	1	2	3
25	1	3	3	3	4	2	1	3.16228
7	1	3	3	2	4	2	1	3.31662
5	1	3	2	2	4	2	1	3.31662
14	1	3	3	2	4	3	1	3.46410
6	1	3	2	2	4	3	1	3.46410
32	1	3	2	2	4	3	1	3.60555
33	1	3	2	2	4	3	1	3.60555
37	2	2	3	3	4	1	2	3.60555
42	1	3	2	2	4	3	1	3.60555
41	1	1	1	1	2	2	1	3.74166
28	1	1	1	1	2	3	1	3.74166
31	1	1	1	1	2	3	1	3.74166
4	2	1	1	1	2	1	2	3.74166
10	2	1	1	1	2	1	2	3.74166
13	2	1	1	1	2	1	2	3.74166
21	2	1	1	1	2	1	2	3.87298
24	2	1	1	1	2	1	2	3.87298

- c. *Euclidean* distance , then the K value will be determined to find the nearest neighbor. In determining the value of K, the authors consider that the number of nearest neighbors from each class is not the same. Because the classification process uses an odd K value. Determining the value of K also depends on the number of classes, that is, if the number of classes is even, the K value should be odd, and vice versa, if the number of classes is odd, then the K value should be even.

In this study the authors determined the odd K value, namely 25 because the training data was an even number, namely 50. Determining the value of K = 25 because the K value is the best K value with the least error than the other K values. From the 1st test data and determining the value of K = 25, the following labels will be generated:

Table 13. Result of Total Data of Each Label

Cat Type	Label	Number of Labels
Angora	1	18
Village	2	7

From the data above, it can be concluded that the results of the 1st data test using the K-Nearest Neighbor method show that the highest number of labels is 1, namely Angora Cat with a total of 18 labels.

3.3. Results Analysis

After testing with 50 training data and 50 test data, the following data will be generated:

Table 14. Test Results With K=25

Data	Body shape	Feather Length	Nose Width	Nose Height	Food Type	Feather Type	tags 1	tags 2	Type of Cat (k-NN)	Type of Cat (Manual)	Test results
1	2	3	3	3	1	2	18	7	1 Angora	2 Village	WRONG

2	1	3	3	3	1	2	18	7	1	Angora	1	Angora	CORRECT
3	2	2	3	3	3	1	10	15	2	Village	2	Village	CORRECT
4	2	1	1	1	2	3	12	13	2	Village	1	Angora	WRONG
5	2	2	2	2	4	2	9	16	2	Village	2	Village	CORRECT
6	2	2	3	3	3	1	10	15	2	Village	2	Village	CORRECT
7	1	3	3	3	1	2	18	7	1	Angora	1	Angora	CORRECT
8	1	3	3	3	1	3	20	5	1	Angora	1	Angora	CORRECT
9	1	1	1	1	2	3	16	9	1	Angora	1	Angora	CORRECT
10	2	2	2	2	4	3	10	15	2	Village	2	Village	CORRECT
11	2	2	3	3	3	1	10	15	2	Village	2	Village	CORRECT
12	1	3	3	3	1	3	20	5	1	Angora	1	Angora	CORRECT
13	2	2	2	2	4	3	10	15	2	Village	2	Village	CORRECT
14	2	2	2	2	4	2	9	16	2	Village	2	Village	CORRECT
15	1	3	3	3	1	2	18	7	1	Angora	1	Angora	CORRECT
16	1	3	3	3	1	3	20	5	1	Angora	1	Angora	CORRECT
17	1	3	3	3	1	3	20	5	1	Angora	1	Angora	CORRECT
18	2	1	1	1	2	2	6	19	2	Village	2	Village	CORRECT
19	1	3	3	3	1	3	20	5	1	Angora	1	Angora	CORRECT
20	2	2	3	3	3	1	10	15	2	Village	2	Village	CORRECT
21	1	3	3	3	1	2	18	7	1	Angora	1	Angora	CORRECT
22	2	2	2	2	4	2	9	16	2	Village	2	Village	CORRECT
23	1	1	1	1	2	3	16	9	1	Angora	1	Angora	CORRECT
24	2	2	2	2	4	2	9	16	2	Village	2	Village	CORRECT
25	1	3	3	3	1	2	18	7	1	Angora	1	Angora	CORRECT
26	1	3	3	3	1	3	20	5	1	Angora	1	Angora	CORRECT
27	1	3	3	3	1	3	20	5	1	Angora	1	Angora	CORRECT
28	2	2	3	3	3	1	10	15	2	Village	2	Village	CORRECT
29	2	2	3	3	3	1	10	15	2	Village	2	Village	CORRECT
30	2	2	3	3	4	1	10	15	2	Village	2	Village	CORRECT
31	1	3	3	3	1	3	20	5	1	Angora	1	Angora	CORRECT
32	2	2	2	2	4	3	10	15	2	Village	2	Village	CORRECT
33	1	1	1	1	2	3	16	9	1	Angora	1	Angora	CORRECT
34	2	2	3	3	3	1	10	15	2	Village	2	Village	CORRECT
35	1	3	3	3	1	2	18	7	1	Angora	1	Angora	CORRECT
36	1	3	3	3	1	3	20	5	1	Angora	1	Angora	CORRECT
37	1	3	3	3	1	3	20	5	1	Angora	1	Angora	CORRECT
38	2	2	3	3	3	1	10	15	2	Village	2	Village	CORRECT
39	1	3	3	3	1	3	20	5	1	Angora	1	Angora	CORRECT
40	1	1	1	1	2	1	6	19	2	Village	1	Angora	WRONG
41	2	2	3	3	3	1	10	15	2	Village	2	Village	CORRECT
42	2	2	3	3	3	1	10	15	2	Village	2	Village	CORRECT
43	1	3	3	3	1	2	18	7	1	Angora	1	Angora	CORRECT
44	2	1	1	1	2	2	6	19	2	Village	2	Village	CORRECT
45	1	3	3	2	1	2	19	6	1	Angora	1	Angora	CORRECT
46	1	3	3	3	1	3	20	5	1	Angora	1	Angora	CORRECT
47	1	3	3	3	1	3	20	5	1	Angora	1	Angora	CORRECT
48	2	2	3	3	3	1	10	15	2	Village	2	Village	CORRECT
49	2	2	3	3	3	1	10	15	2	Village	2	Village	CORRECT
50	2	2	3	3	3	1	10	15	2	Village	2	Village	CORRECT

After testing the data and comparing the test results with the *K-Nearest Neighbor method* and the manual data obtained. It can be seen that there are 3 data that do not match or are wrong. So that the level of accuracy of the test can be determined by calculating the following formula:

$$\begin{aligned} \text{Accuracy rate} &= \frac{47}{50} \times 100\% \\ &= 94\% \end{aligned}$$

So from testing the determination of the type of cat using the *K-Nearest Neighbor method* with 50 training data and 50 testing data, the accuracy rate is 94% and the error rate is 6%.

4. CONCLUSION

This study compares the classification of cat breeds with the *K-Nearest Neighbor method* and manual determination (directly in the field). The study was conducted using 50 training data and 50 testing data. With 6 parameters taken, body shape, hair type, hair length, nose height, nose width, and type of food. After testing with the *K-Nearest Neighbor method*, there are 3 incorrect data (does not match the manual data) and 47 correct data. So that it can be seen the level of accuracy of cat type classification using the *K-Nearest Neighbor method* is 94% with an error rate of 6%. It can be seen that the results of research using the *K-Nearest Neighbor method* are very close to direct research in the field.

REFERENCES

- [1] Muhammad Afif A. F, et al. *Cat Race Classification using Convolutional Neural Network (CNN) Algorithm*. E-Proceeding of Engineering. 2021; Vol.8, p. 2
- [2] Anggi N. P, Isnawati. *Morphogenetics of the House Cat (Felis domesticus) as a Natural Predator Breeding Tool for Rodents*. Bio Lantern. 2022; Vol.11, p. 217
- [3] Hasmawati, Nangi, et al. *Goods Sales Prediction Application Using the k-Nearest Neighbor (k-NN) Method (Case Study of Tumaka Mart)*. Semantics. 2017. Vol.3, no.2, pp. 151-160
- [4] Ahmad Khairi, et al. *Implementation of K-Nearest Neighbor (KNN) for Classification of Pre-Prosperous Communities in Sapikerep Village, Sukapura District*. Trilogy: Journal of Technology, Health, and Humanities. 2021
- [5] Muhammad Ihsan Al-Hafiz . *Implementation of the K-Nearest Neighbor (K-NN) Algorithm in Determining Cat Types*. Thesis. Palangkaraya & College of Information and Computer Management (STMIK) Palangkaraya. 2021.
- [6] Permana Putra, et al. *Analysis of the K-Nearest Neighbor (KNN) Method in Iris Flower Data Classification*. Kaputama Informatics Engineering Journal (JTIK). 2022; Vol.6
- [7] Anton Yudhana, et al. *K-NN Algorithm with Euclidean Distance for Predicting the Yield of Sengon Sawmills*. 2020. TRANSMISSION
- [8] HAF Willmen TB Panjaitan, et al. *Rice Harvest Prediction Using the k-Nearest Neighbor Algorithm*. Pros. 5th Year SNATIVE. 2018. Pp. 621-628
- [9] M.S. Mustafa and IW Simpen . *Designing a Timely Graduation Prediction Application for New Students with Data Mining Techniques (Case Study: Student Academic Data at STIMIK Dipanegara Makassar)*. Creat, Inf, Technol, J. 2015. Vol.1, no.4, pp.270-281
- [10] A. Hamdi, et al. *The Timeseries K-Nearest Neighbor Regression Method in Predicting Outgoing Goods at the Warehouse of PT Putra Preneur Banjarbaru*. Monday. Nas. Computing Science. 2019. Vol.2, pp.37-45
- [11] I. Drajana and C. Rally. *Prediction of Cococnut Oil Production Using k-Nearest Neighbor and Backward Elimination*. TECNOSCIENZA. 2018. Vol.13, no.1, pp. 51-64
- [12] N. Lizarti and AN Ulfah. *Application of the k-Nearest Neighbor Algorithm for Determining Study Specialization (Case Study: STIMIK AMIK Riau Informatics Engineering Study Program)*. Fountain Informatics J. 2019. Vol.4, p.1-7
- [13] I. Riadi, et al. *Handwriting Identification of Japanese Katakana Letters with the Euclidean Method*. J-SAKTI (Journal of Computing and Information Science). 2020. Vol. 4, no. 1, p. 29

- [14] A. Fadlil and Saifudin. *Wood Image Identification System Based on Texture Using Gray Level Cooccurrence Matrix (GLCM) with Euclidean Distance Classification*. 2015. Vol.19, no.3, pp.181-186
- [15] Sunardi, A. Fadlil, and Suprianto. *Sentiment Analysis Using the Naive Bayes Classifier Method in Student Questionnaires*. *Saintekbu*. 2018. Vol.10, no.2, pp. 1-9.