

Impression Classification of Endek (Balinese Fabric) Image Using K-Nearest Neighbors Method

Gede Aditra Pradnyana^{*1}, I Komang Agus Suryantara², I Gede Mahendra Darmawiguna³

^{1,2,3}Universitas Pendidikan Ganesha/Informatics Engineering Education Departement,
gede.aditra@undiksha.ac.id^{*1}, helloaguss@gmail.com²,
mahendra.darmawiguna@undiksha.ac.id³

Abstract

An impression can be interpreted as a psychological feeling toward a product and it plays an important role in decision making. Therefore, the understanding of the data in the domain of impressions will be very useful. This research had the objective of knowing the performance of K-Nearest Neighbors method to classify endek image impression using K-Fold Cross Validation method. The images were taken from 3 locations, namely CV. Artha Dharma, Agung Bali Collection, and Pengrajin Sri Rejeki. To get the image impression was done by consulting with an endek expert named Dr. D.A Tirta Ray, M.Si. The process of data mining was done by using K-Nearest Neighbors Method which was a classification method to a set of data based on learning data that had been classified previously and to classify new objects based on attributes and training samples. K-Fold Cross Validation testing obtained accuracy of 91% with K value in K-Nearest Neighbors of 3, 4, 7, 8.

Keywords: Classification, Endek, Impression, K-Nearest Neighbors

1. Introduction

Bali is a place that has culture and nature tourist attractions that are very famous in Indonesia. Each year tourist visits increased, seen in 2014 with a visit of foreign tourists as many as 3766 638, then in 2015 increased to 4,001,83. And the last record at the time this study was made in 2016 foreign tourists visiting Bali reached 4,927,937, while for domestic tourists the number of tourists visiting Bali in 2014 reached 6,394,307, and increased to 7,147,100 in 2015 [1]. This shows that Bali tourism is still getting an attention of tourists, for tourist destinations that are worth to visit.

Bali has a society with a diverse local culture and also a famous weaving craft namely endek. Endek is a fabric made from weaving process and a special souvenir for foreign and domestic tourists. In general, endek fabrics are made by women who have creative and high art creations. The endek fabric is an applied art, which means art that has benefits for human beings in daily life other than its art value [2]. Bali Provincial Government requires employees to use clothing made from endek fabric every Thursday and Friday as an office uniform. In addition, teachers are required to wear endek clothes as a uniform when teaching.

The uniqueness of endek fabric is its pattern. Endek fabrics have diverse patterns. Some sacred endek fabric patterns that only be used for activities at temples or other religious activities. There are also endek patterns that can only be used by certain people, such as kings or descendants of nobility. The endek product is increasingly varied with the design and pattern that are slightly modern but do not leave the art of aesthetics and a strong culture of the endek itself. In addition, Balinese people are generally loaded with traditional ceremonies and religious rituals of course use endek fabric and growing consumer interest to have the latest endek pattern.

An impression can be interpreted as a psychological feeling toward a product [3]. An impression can affect a person in decision making and selection of a good product. Because impression plays an important role in decision making. The understanding of data in the domain of impressions will be very useful. Endek clothes can be used by men and women. The impression between male and female clothing may vary despite using the same endek fabric. Frequently in the use of endek cloth, users often choose the endek with a wrong impression in a particular event. From the results of interviews with endek craftsmen, usually, the craftsmen do production by order and produce the same pattern of the order to stock in their shops without knowing the impression of the pattern itself.

Studies on Endek and impression have respectively been carried out. A study on impression done [4] was entitled "Generic and Specific Impression of Estimation of Clothing Fabric Images Based on Machine Learning." [4] The purpose of this study was to predict impression of textile images based on human emotion concepts by using Support Vector Regression.

Another study on impression has also been conducted [3] with the title "The Influence of Color Feature on Impression Classification of Indonesian Batik Images Using Probabilistic Neural Network." [3] The purpose of this study was to determine the influence of color features on impression classification of Indonesian Batik Images.

Another study on impression has been done [5] with the title "Study on the Emotion Factor Space of Men's T-Shirt." [5] The purpose of this study was to determine the impressions on men's shirts. The method used was Factor Analysis. Apart from that, a study concerning on impression has also been conducted [6] with the title "Predicting Personalized Image Emotion Perceptions in Social Media." [6] The purpose of this study was to determine the prediction on image impression by combining* a variety of factors (social context, temporary evolution, and neighborhood influence). The method used to gain the intended impression was Rolling Multi-task Hypergraph Learning.

There are many methods of data mining that can be used to discover knowledge. K-Nearest Neighbors (K-NN) is a classification method for data sets based on previously classified data learning and to classify new objects based on attributes and training data [7]. The K-Nearest Neighbors algorithm is a supervised learning that to find new patterns in the data using existing data patterns. The advantages of K-Nearest Neighbors are effective with noisy and large data training.

Previous research related to endek classification titled "Texture Analysis on Image Motif of endek Bali using K-Nearest Neighbors Classification Method" [8] reveals that K-NN method is suitable for endek classification because endek cloth has similarity pattern to each other, so the K-NN method can reduce the noise. The research resulted in an accuracy of 57%. This study is expected to provide information on impression classification of Endek image. From this comprehension, the users of endek fabric can make choices at endek fabrics appropriate with the impression desired. The obstacle in this study included the absence of data set, so that the researcher established an own-version of data set. Another challenge was that it was quite hard to determine the endek expert, for the study on endek especially on data mining field was still rarely performed.

2. Research Method

2.1 Endek

An endek fabric is formed by weaving or crossing two groups of strings upright to each other. The principle of making endek fabric is to cross the pakan string (the length of the endek fabric) in the gap of the lusi string (the width of the endek fabric). An endek fabric is an applied art, which means an art that has benefits for humans in everyday life other than its art value [2].

2.2 Impression

The impression can be interpreted as a psychological feeling toward a product [3]. A piece of endek fabric does not only keep the beauty values of the patterns and blend colors, but it also holds the philosophical value and adaptation strategies of the supporting community. Because there are differences in the value system, symbols, philosophical meanings and adaptation strategies, the impression of the endek fabric created by the community is different from others. Impressions were used in this study using impression on research conducted by [3] which amounted to 6 with descriptions by KBBI shown in Table 1.

Table 1. Impression

Impression	Description
Adult	Not a child or a teenager anymore.
Calm	Quiet; Relaxed:
Warm	Happy, joy
Dynamic	Full of spirit and energy
Masculine	Male
Feminine	Female

2.3 Digital Image Processing

Literally, an image is an image in a two-dimensional field and if viewed from a mathematical point of view, the image is a continuous function of the intensity of light in a two-dimensional field [9]. An images can be grouped into two parts: the still image of a single fixed image and a moving image that is a series of still images displayed in a row, thus giving the impression on the eyes as moving images [9]. Each image in a row is called a frame. Images that appear on a movie or television are made up of thousands to hundreds of thousands of frames.

The digital image is an image composed of discrete pixels of the quantized brightness and color [9]. Basically, an image that has a continuous color and brightness needs to be changed in the form of color information, brightness, etc. to a discrete value in order to become a digital image.

2.4 Feature Extraction

An image feature is a function of one or more measurements, each of which determines the measured property of an object, and is calculated in such a way as to determine the characteristics of the image [10]. Mathematically, each feature extraction is an encoding of the n dimensional vector called the feature vector. The feature vector component is calculated by image process and analytical techniques and is used to compare images with one another.

Texture is a characteristic possessed by a region large enough so that naturally these traits can be repeated in the area or the regularity of certain patterns formed from the arrangement of pixels in the image. And if the extent is enlarged without changing the scale, the nature - the surface properties of the expansion result are of a similar nature to the original surface [11]. Texture analysis aims at identifying the appropriate parameters with the characteristics of the objects in an image.

Gray Level Co-occurrence Matrix (GLCM) was first proposed by Haralick in 1979 with 28 features to explain spatial patterns. GLCM is a texture feature extraction technique that uses a statistical approach in which there are a number of pixel pairs for a certain level of brightness separated by distance and slope between pixel [10]. Feature extraction obtained by using GLCM method is contrast, correlation, homogeneity, and energy.

The color feature extraction can be done by the mean which is the extraction of the color feature on the image can be done by using the mean on each RGB channel and the standard deviation, which the color feature extraction in the image can be done by using standard deviation on each RGB channel

2.5 Data Mining

Data mining is a process of finding a correlation, new patterns, and new trends useful in big data using pattern recognition, such as statistical and mathematical techniques [12]. Data mining is used to automatically explore thoroughly and describe complex relationships on very large datasets. The dataset is a tabular data set, as is widely implemented in relational database management. However, data mining techniques can also be applied to other data representations, such as spatial data domains, text-based, and multimedia (images).

The process of data mining will extract valuable information by analyzing the existence of patterns or relationships of certain relationships of large data. Automated analysis performed by data mining exceeds that traditional decision support systems that have been widely used. Data mining can perform the information retrieval process, predicting in large databases.

2.6 K-Nearest Neighbors

Nearest Neighbors (NN) is one of the most popular methods of classification and a lazy learner classification because it delays the training process until there is a test data that already classify, and the new method will run its algorithm [13]. Learning process will increase the system knowledge. Hence lazy learner can represent complex knowledge models that other algorithms may not be able to do.

K-Nearest Neighbors is one of the Nearest Neighbors-based methods. K-Nearest Neighbors (K-NN) is a classification method for data sets based on previously classified data learning and for classifying new objects based on attributes and training samples [7]. The K-NN algorithm works based on the shortest distance from the query instance to the test data to determine its K-NN. One way to calculate the distance of a neighbor is by using the Euclidian

distance method. Euclidean distance between two points. If $X_1 = (x_{11}, x_{12}, \dots, x_{1n})$ and $X_2 = (x_{21}, x_{22}, \dots, x_{2n})$, it can be expressed by Equation 1

$$distance(X_1, X_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \tag{1}$$

2.7 K-Fold Cross Validation

Cross Validation is a common method used to evaluate classifier performance. In a cross-validation approach, each tuple is used multiple times in the same amount of training data and is appropriate for test data. For accuracy calculation the following Equation 2, where a is the number of correctly classified data, and b is the amount of data classified incorrectly.

$$accuracy = \frac{a}{a + b} \tag{2}$$

2.8 Research Overview

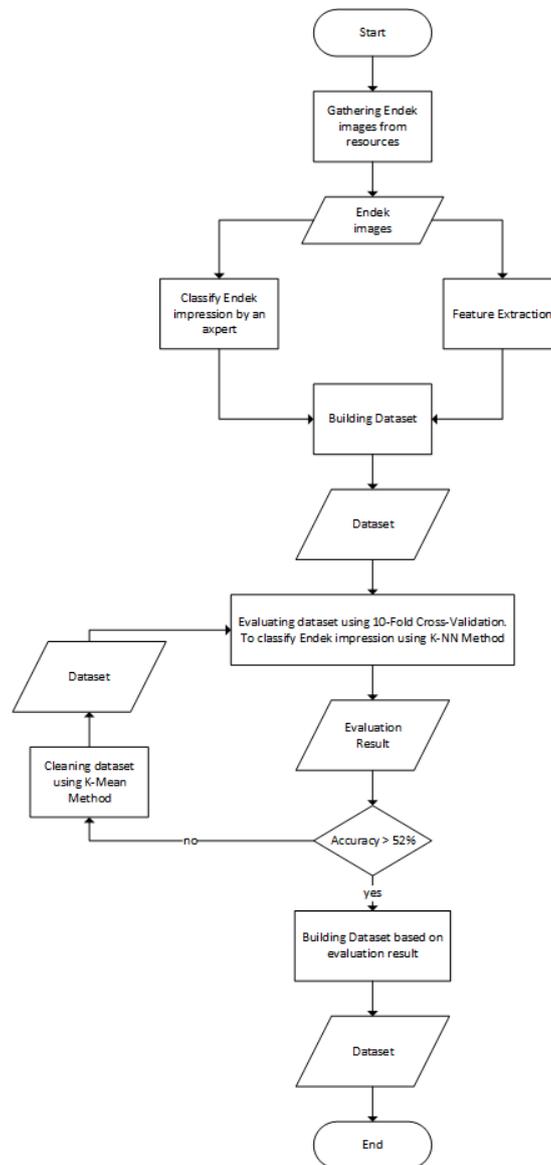


Figure 1. Endek Dataset Development and Impression Classification using K-NN

Figure 1 shows how the stages of creating datasets and the development of methods to classify endek image impressions. The earliest stage that was carried out was the collection and production of data sets from the original endek fabric. This is done considering that when this research was conducted, data sets related to endek images did not yet exist. After the dataset is complete, the next step is to apply the k-NN method to image classification. Besides testing the accuracy of the impression classification process, this study also focuses on producing good endek image datasets. Therefore, the accuracy of the endek impression classification with the K-NN method is used as the basis for fixing the dataset. The improvement of endek image datasets is done by applying the K-Means clustering method, to label or re-categorize existing image datasets. A more detailed explanation of the results of this stage can be seen in the result and analysis section.

3. Result and Analysis

3.1 Data Source

Figure 2 in this research, the data used is Endek images taken with RAW format. The advantage of using images with RAW format is the uncompressed image file, thus storing more information when compared to the compressed image file. Images were taken at three locations in Bali, namely CV. Artha Dharma located in Buleleng, Agung Bali Collection located in Bangli, and Pengrajin Sri Rejeki located in Klungkung.



Figure 2. Endek

3.2 Dataset

The process building training data was done by collecting endek image which then each image was given an impression. After the process, it performed a preprocessing data to maximize the training data that is Data Cleaning is a process of cleaning data is done that eliminates the endek image that has the same pattern and color, Data Integration is a process of combining data from 3 craftsmen in Buleleng, Bangli, and Klungkung and the impression given earlier. Data Selection is a process to choose the endek that has negative impression or not in accordance with the specified impression. Data Transformation is a process that is done to crop on each endek to the same size that is 1500px x 1500px, using the uncompressed image format, named TIF. Data set established in the respective impressions is as follows Table 2.

Table 2. Number of Data

No.	Impression	Number of Data
1	Adult	29
2	Calm	15
3	Warm	20
4	Dynamic	18
5	Masculine	11
6	Feminine	17

3.3 Impression Classification using K-Nearest Neighbors Method

Data mining process aimed at determining the impressions of a new endek image that were not previously classified the impression using the K-Nearest Neighbors method. This test used 10 training data in Table 3 and one testing datum in Table 4.

Table 3. Data Training

1	2	3	4	5	6	7	8	9	10	11
0,51	0,88	0,18	0,84	119,86	104,23	120,02	49,94	47,09	36,16	Feminine
0,15	0,96	0,19	0,93	148,22	103,34	107,81	37,61	50,21	37,42	Feminine
0,30	0,90	0,17	0,87	216,45	119,35	172,17	25,56	48,63	33,26	Feminine
0,05	0,95	0,38	0,97	216,99	127,99	39,40	26,33	22,73	17,95	Adult
0,06	0,86	0,62	0,97	107,40	158,89	194,01	26,37	16,43	19,93	Warn
0,06	0,96	0,35	0,97	96,21	129,20	111,74	25,71	26,46	25,93	Calm
0,10	0,96	0,23	0,95	86,51	77,09	113,95	51,25	36,55	28,05	Dynamic
0,65	0,57	0,17	0,73	147,50	144,57	148,11	32,06	25,28	20,44	Masculine
0,05	0,92	0,55	0,97	109,88	150,99	163,92	23,37	16,38	17,64	Dynamic
0,15	0,93	0,24	0,93	115,49	97,76	176,68	35,05	35,22	15,77	Warm

1. Contrast; 2. Correlation; 3. Energy; 4. Homogeinity; 5. Mean in Red Channel; 6. Mean in Green Channel; 7. Mean in Red Channel; 7. Standard Deviation in Red Channel; 8. Standard Deviation in Green Channel; Standard Deviation in Blue Channel;

Table 4. Data Testing

1	2	3	4	5	6	7	8	9	10
0,48	0,81	0,14	0,79	151,09	89,40	141,04	37,69	35,93	27,22

1. Contrast; 2. Correlation; 3. Energy; 4. Homogeinity; 5. Mean in Red Channel; 6. Mean in Green Channel; 7. Mean in Red Channel; 7. Standard Deviation in Red Channel; 8. Standard Deviation in Green Channel; Standard Deviation in Blue Channel;

The calculation of K-Nearest Neighbors will be performed using Table 3 as training data and Table 3 as testing data. Determine the parameter K, that is a number of nearest neighbors. In this case, the use of K is equal to 4. Calculate the distance between test data and all training data y. To calculate the distance between the train data and the test data it was used Euclidean distance. Using Equation 3, we get the distance table between the training data and the test data.

$$\begin{aligned}
 \text{distance}(X_1, X_2) &= \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \\
 &= \sqrt{(0.153 - 0.067)^2 + (0.892 - 0.800)^2 + (0.258 - 0.276)^2 + (0.924 - 0.928)^2 + (241.529 - 6.864)^2 + (24.718 - 38.867)^2 + (248.427 - 217.404)^2 + (40.762 - 93.437)^2 + (66.558 - 197.486)^2 + (193.851 - 67.702)^2} \\
 &= \sqrt{1991.518} = 44,626
 \end{aligned} \tag{3}$$

From all results of distance between testing data with training data, they were then sorted by value of Euclidian Distance to get Euclidian Distance table at Table 5.

Table 5. Euclidian Distance

No	Rank	Euclidean distance	Class
1	2	44,62642665	Feminine
2	1	40,18183377	Dynamic
3	7	80,517449	Feminine
4	10	128,6550405	Masculine
5	9	100,5153241	Warm
6	6	75,42568524	Warm
7	5	72,39663089	Dynamic
8	4	57,41835539	Feminine
9	8	81,81087377	Calm
10	3	52,39845696	Adult

Selecting data with the closest Euclidean distance, because from using K equal to 4 the data selected were the data that were the top 4 in the Euclidian Distance table. To determine the impression on the testing data it was used the majority, so the impression obtained from the testing data was Feminine.

3.4 K-Fold Cross Validation

K-Fold Cross Validation is a method to validate the accuracy of a data mining algorithm based on a particular data set. This study used 10-Fold Cross Validation. This test started from partitioning the data set into 10 partitions which later on 1 part of the partition was used as testing data and the rest was used as training data. The process was done by each partition of the data set that became test data on each iteration.

The K-Fold Cross Validation testing process produced an accuracy of 40% in K for K-NN is 10. The accuracy of train data was small due to the unbalanced amount of data in each class, so the training data had to be processed again to get better accuracy.

To get a better accuracy data was done data cleaning. The first step was cluster the data into 6 clusters using K-Means method. It chose the clustering results that each cluster has a number of members is almost equal to each other. After the clustering results were obtained, the next step was to determine the impression on each cluster group, by choosing the most dominant impression in each cluster group. In the process of determining impressions, there was one cluster that could not determine the impression, so the data from that cluster was removed from data training, and it eliminated one impression namely Dynamic. After the training data were completed, the accuracy test was done using K-Fold Cross Validation method. The results of this test indicated that the accuracy increased significantly after the training data was re-processed that was 91% in K for K-NN was a 3,4,7,8.

4. Conclusion

The conclusion basically referred to the purpose of research that had been designed previously. Success or failure in achieving these objectives was evidenced in the implementation and testing stages. Based on the research that has been done, it can draw the conclusion (1) the application of endek image impression classification using K-Nearest Neighbors method has succeeded in classifying the endek image impression into six classes, namely adult, calm, warm, dynamic, masculine, and feminine, by using the results of design analysis of the classification model that has been proven by testing using a white box and black box method. (2) The designed application has the highest average accuracy rate of 40% on the K-Fold Cross Validation test with K in K-NN is equal to 10. This level of accuracy is still low due to the imbalance amount of data in each class. However, after the training data is processed again, it produces a higher accuracy rate of 90%, and eliminating one dynamic impression.

References

- [1] Dinas Pariwisata Provinsi Bali, "Statistik." [Online]. Available: <http://www.disparda.baliprov.go.id/id/Statistik3>. [Accessed: 08-Mar-2017].
- [2] Yayat, "Pengertian Seni Rupa Terapan beserta Contohnya," 2015. [Online]. Available: <http://www.pakmono.com/2015/10/pengertian-seni-rupa-terapan-beserta-contohnya.html>. [Accessed: 18-Feb-2017].
- [3] Agung Nilogiri, "Pengaruh Fitur Warna pada Klasifikasi Impresi Citra Batik Indonesia Menggunakan Probabilistic Neural Network," *Jurnal Sistem dan Teknologi Informasi Indoensia*, Vol. 01, No. 01, Pp. 57–63, 2016.
- [4] X. Chen, Y. W., Chen, D., Han, X. H., & Huang, "Generic and Specific Impression Estimation of Clothing Fabric Images Based on Machine Learning," in *12th International Conference on Fuzzy System and Knowledge Discovery (FSKD)*, 2016.
- [5] Z. C. Z. Hai-bo, "Study on the Emotion Factor Space of Men's T-Shirt," in *3rd International Conference on Information Science and Control Engineering*, 2016.
- [6] S. Zhao, H. Yao, Y. Gao, and G. Ding, "Predicting Personalized Image Emotion Perceptions in Social Networks," *IEEE Transaction Affective Computing*, Vol. 10, No. X, Pp. 1–14, 2016.
- [7] A. W. K. Yodha, Johannes Widagdo, "Pengenalan Motif Batik Menggunakan Deteksi Tepi Canny dan K-Nearest Neighbor," *Techno.COM*, Vol. 13, No. 4, Pp. 2014, 2014.
- [8] I. G. S. Rahayuda, "Texture Analysis on Image Motif of Endek Bali using K-Nearest Neighbor Classification Method," *International Journal of Advance Computer Science and*

- Applications, Vol. 6, No. 9, Pp. 205–211, 2015.
- [9] R. Munir, "*Pengolahan Citra Digital*," Bandung: Informatika, 2004.
- [10] I. Witten, E. Frank, M. Hall, and C. Pal, "*Data Mining: Practical Machine Learning Tools and Techniques*." Second Edition, San Francisco: Diane Cerra, 2005.
- [11] M. Yuda Permadi, "*Aplikasi Pengolahan Citra untuk Identifikasi Kematangan Mentimun Berdasarkan Tekstur Kulit Buah Menggunakan Metode Ekstraksi Ciri Statistik*," Jurnal Informatika, Vol. 9, No. 1, Pp. 1028–1038, 2015.
- [12] A. A. M. Nixon, "*Feature Extraction and Image Processing*," London: Elsevier. Ltd, 2008.
- [13] J. Han and M. Kamber, "*Data mining: Concepts and Techniques*," Amerika Serikat: Diane Cerra, 2006.