JMTAS: Journal of Multimedia Technology and Applied Software
Published: 30 April 2025
E-ISSN: 3047 - 4760

Analysis of the Level of Satisfaction of Using the Majalengka Digital Samsat Website Using the K-Nearest Neighbor Algorithm.

[1st] Muhammad Abdul Jabbar 1*, [2nd] Fikri Akbar Falcata W 2*, [3rd] Rifqi Syamsul Fuadi 3*, [4rd] Ade Taopik Hidayatuloh 4*.
[1st] Teknologi Rekayasa Multimedia 1, Politeknik Mardira Indonesia 1, [2nd] Teknologi Rekayasa Multimedia 2, Politeknik Mardira Indonesia 2,
[3rd] Teknologi Rekayasa Multimedia 3, Politeknik Mardira Indonesia 3, [3rd]Manajemen Transportasi Laut, Politeknik Pelayaran Banten 4
Email : jabbarmuhammad20@gmail.com 1*, fikriakbar094@gmail.com 2*, rifqi@poltekmi.ac.id 3*, adetaufik61@gmail.com 4.

The development of mobile applications is currently very rapid. Many applications have emerged with the aim of meeting various human needs, ranging from hobbies, lifestyle, household needs, to religious aspects and others. However, because of the many applications with similar functions, competition between developers is inevitable. They compete to attract user interest by presenting superior features that differentiate their applications from competitors. In this context, sentiment analysis is one of the important methods for understanding how users perceive an application. The K Nearest Neighbors (KNN) method was chosen as the sentiment analysis approach due to its simple, intuitive nature, and its classification capabilities based on classification.

Keywords: Mobile applications; Sentiment Analysis; Data mining; K-Nearest Neighbor method;

Perkembangan aplikasi mobile saat ini berlangsung sangat pesat. Banyak aplikasi bermunculan dengan tujuan untuk memenuhi berbagai kebutuhan manusia, mulai dari hobi, gaya hidup, keperluan rumah tangga, hingga aspek keagamaan dan lainnya. Namun, karena banyaknya aplikasi dengan fungsi serupa, persaingan antar pengembang pun tidak terhindarkan. Mereka berlomba-lomba menarik minat pengguna dengan menghadirkan fitur-fitur unggulan yang membedakan aplikasi mereka dari kompetitor. Dalam konteks ini, analisis sentimen menjadi salah satu metode yang penting untuk memahami bagaimana persepsi pengguna terhadap suatu aplikasi, Metode K-Nearest Neighbors (K-NN) dipilih sebagai pendekatan analisis sentimen karena sifatnya yang sederhana, intuitif, serta kemampuan klasifikasinya berdasarkan klasifikasi.

Kata Kunci: Aplikasi mobil,; Sentiment Analysis; Data mining; Metode K-Nearest Neighbor.

INTRODUCTION

The development of mobile applications is currently very rapid. Various applications have been created to meet various human needs, ranging from hobbies, lifestyle, household matters, religion, to other needs. In its implementation, many applications offer similar functions, thus creating competition between developers to attract as many users as possible. One way to do this is by presenting superior features that differentiate their applications from competitors.

In the development process, user input is very important. Comments and reviews left by users on platforms such as the App Store can be a valuable source of information. These reviews often reflect the level of user satisfaction after using the application. The criticism and suggestions written can be used by developers to improve and enhance the quality of their applications.

For this reason, a classification process is needed for these comments to determine whether the content is positive or negative. By understanding user sentiment, developers can make more appropriate decisions in future application development to better suit user needs and expectations.

Several empirical studies illustrate the use of the K-Nearest Neighbors (K-NN) algorithm for sentiment analysis in this context:

a. Nadya Bethry Balqies Tjikdaphia and Sulastri (2023) conducted a comparative study of sentiment classification using Naive Bayes, K-NN, and SVM on Mobile JKN app reviews. With a 90% training and 10% testing split, K-NN achieved 75% accuracy, Naive Bayes reached 87%, while SVM led at 95%
b. Siti Rihastuti, Afnan Rosyidi, Handoko, and M. Nur Juniadi (2024) applied K-NN for sentiment analysis on 1,000 scraped user reviews of the JKN Mobile app. Using K = 11 and an 80:20 train-test ratio, they achieved 89% accuracy, 97% precision, and 86% recall

METHODOLOGY

1. THIS RESEARCH WAS CONDUCTED IN SEVERAL STAGES, NAMELY:

a. Data Collection Techniques
b. Data collection techniques used in this final assignment research are:
c. Field Study Field study is intended to obtain data and information directly. The data collection techniques used are:

d. Observation, namely the collection of data and information carried out by directly observing the research object.
e. Interview, namely data collection by conducting questions and answers with related parties such as parents.
f. Literature Study Data is obtained through literature books related to the problem to be studied, namely the problem of adolescent psychology as reference material for the author.

2. KNN (K-NEAREST NEIGHBOR)

The K-Nearest Neighbor (KNN) algorithm works by utilizing the proximity of data as the basis for determining the classification or prediction of new data (query instance). Several previous studies have discussed the application of classification methods in data mining. One such study was conducted by Edi Permadi Budiyono, Nerfita Nikentari, ST., M.Cs, and Sulfikar Sallu, S.Kom., M.Kom (2014), titled Analysis of Gold Karat Classification Using the K-Nearest Neighbors (KNN) Method. This study showed that the KNN algorithm was able to achieve an accuracy rate of up to 92.93% in classifying gold karat levels.

Another reference comes from a study by Ricky Imanuel Ndaumanu, Kusrini, and M. Rudyanto Arief (2014), which examined Prediction Analysis of Student Dropout Rates Using the K-Nearest Neighbor Method. Although the predictive system in this study did not yield maximum accuracy, it was still considered reasonably feasible for use. However, the researchers emphasized that in order to obtain optimal validation results, it is necessary to have a balanced dataset between active students and those who have dropped out.

Based on these studies, the present research employs the K-Nearest Neighbor algorithm to analyze and measure the accuracy level in classifying user satisfaction with an application.

RESULT AND DISCUSSION

1. RESEARCH FRAMEWORK

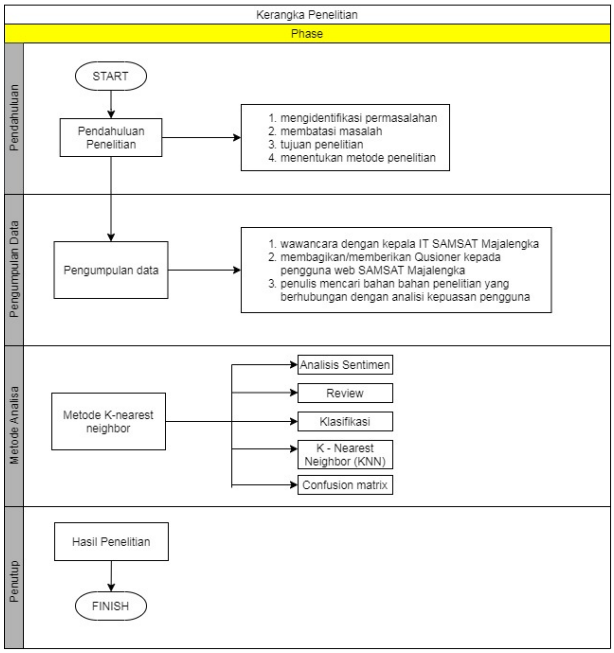

Figure 1. Research Framework

2. K-NEAREST NEIGHBOR

K-Nearest Neighbor is often used in classification with the purpose of this algorithm is to classify new objects based on attributes and training samples. The K-Nearest Neighbor (K-NN) algorithm is a method for classifying objects based on learning data that is closest to the object. This technique is very simple and easy to implement. Learning data is projected into a multidimensional space, where each dimension represents a feature of the data. This space is divided into parts based on the classification of the learning data. A point in this space is marked with class c if class c is the most common classification found in the k nearest neighbors of the point. The proximity or distance of neighbors is usually calculated based on the Euclidean distance. To define the distance between two points, namely the point on the training data (x) and the point on the testing data (y), the Euclidean formula is used as follows: $similarity(T,S)$ =

Description:

T: New case

S: Cases in storage

N: Number of attributes in each case: individual attributes between 1 to n

F: Similarity function of attribute I between case T and case S

W: Weight given to attribute i The following is a table of the range of public satisfaction with the SAMSAT Majalengka service.

$$\frac{\sum_{i=0}^{n} f(T_i,S_i)*W_i}{W_i}$$

Table 1. Combined Value of User Satisfaction Level

number of questionnaires	Categories
<50-26	satisfied
<25-0	not satisfied

The questionnaire consists of 50 sheets, and there are two categories, consisting of SATISFIED, NOT SATISFIED. So to calculate the combined value of the level of satisfaction, the number of questionnaires is divided by the number of categories.

Tabel 2. Gender Category

Gender	Categories
<50-26	Man
<25-0	Woman

The questionnaire consists of 50 sheets, and there are two types of gender categories, consisting of Male, Female. So to calculate the gender category of satisfaction level, the number of questionnaires is divided by the number of categories. Figures 4.3 and 4.4 visualize data in the form of pie and bar datasets, where the pie chart shows a comparison of the total number of datasets and the bar chart shows the division of Gender Category data.

The formula used to calculate the proximity between the two lists of taxpayers, so that the results of taxpayers who are satisfied and dissatisfied with the SAMSAT Majalengka service are as follows:

$similarity(T,S) = \frac{\sum_{i=0}^{n} f(T_i, S_i) * W_i}{W_i}$

Description:

T: New case

S: Cases in storage: number of attributes in each case

F: Similarity function of attribute I between case T and case S

W: Weight given to attribute i Proximity is usually between 0 and 1. A value of 0 means that the two cases are absolutely not similar, conversely for a value of 1 the case is absolutely similar. The weight between one attribute and another on the non-primary objective attribute is defined with a different value.

3. KNN ALGORITHM TESTING

3.1. CONFUSION MATRIX

In this test, testing data with a K value range of 1-50 is used. The formula used is as follows:

(Satisfied + Satisfied) / (Satisfied + Dissatisfied + Dissatisfied + Satisfied)

K value = 1.

Based on the table above, the following percentage values are obtained:

1. The lowest accuracy is at K = 12 with a percentage of 58% and the highest is at K = 10 with a percentage of 80% for the "Satisfied" classification.
2. The lowest accuracy is at K = 10 with a percentage of 80% and the highest is at K = 12 with a percentage of 58% for the "Dissatisfied" classification.
3. The lowest accuracy is at K = 12 with a percentage of 58% and the highest is at K = 10 with a presentation of 80% for the classification of the combination of "Satisfied" and "Dissatisfied".

3.2. IMPLEMENTATION OF THE BEST MODEL WITH TEST DATA

The test data used is questionnaire data with 50 respondents. While the K value used is: statements from the questionnaire.

1. K value = 1 The result obtained is the number of respondents who have a level of satisfaction of 38 respondents and are dissatisfied by 12 respondents.
2. K value = 2 The result obtained is the number of respondents who have a level of satisfaction of 35 respondents and are dissatisfied by 15 respondents.
3. K value = 3 The result obtained is the number of respondents who have a level of satisfaction of 35 respondents and are dissatisfied by 15 respondents.
4. K value = 4 The result obtained is the number of respondents who have a level of satisfaction of 35 respondents and are dissatisfied by 15 respondents.
5. K value = 5 The result obtained is the number of respondents who have a level of satisfaction of 33 respondents and are dissatisfied by 17 respondents.
6. K value = 6 The result obtained is the number of respondents who have a level of satisfaction of 30 respondents and are dissatisfied by 20 respondents.
7. K value = 7 The result obtained is the number of respondents who have a level of satisfaction of 30 respondents and are dissatisfied by 20 respondents.
8. K value = 8 The result obtained is the number of respondents who have a level of satisfaction of 34 respondents and are dissatisfied by 16 respondents.
9. K value = 9 The result obtained is the number of respondents who have a level of satisfaction of 38 respondents and are dissatisfied by 12 respondents.
10. K value = 10 The result obtained is the number of respondents who have a level of satisfaction of 40 respondents and are dissatisfied by 10 respondents.
11. K value = 11 The result obtained is the number of respondents who have a level of satisfaction of 37 respondents and are dissatisfied by 13 respondents.
12. K value = 12 The result obtained is the number of respondents who have a level of satisfaction of 29 respondents and are dissatisfied by 21 respondents.
13. K value = 13 The results obtained are the number of respondents who have a level of satisfaction of 33 respondents and those who are dissatisfied are 17 respondents.

3.3. CALCULATING EUCLIDEAN DISTANCE

In the calculation to find the Euclidean value, enter the data into the equation.

Where: x2i is the test data and x1i is the training data as much as p (attributes, namely: number of respondents, number of satisfied and number of dissatisfied)

$$De(1) = \sqrt{50^2 + (38 - 12)^2}$$
$$= \sqrt{2500 + 676}$$
$$= \sqrt{3176}$$
$$= 56{,}3560$$

$$d_e = \sqrt{\sum_{1-i}^{p} (x_{2i} - x_{1i})^2}$$

Perform the same calculation for all training data Euclidean calculation results for all training data.

Table 3. Euclidean Calculation Results

No	Sum Euclidean	Euclidean
1.	3176	56,3560
2.	2900	53,8516
3.	2900	53,8516
4.	2900	53,8516
5.	2756	52,4976
6.	2600	50,9901
7.	2600	50,9901
8.	2824	53,1413
9.	3176	56,3560
10.	3400	58,3095
11.	3076	55,4616
12.	2564	50,6359
13.	2756	52,4976

3.4. CALCULATING WEIGHTING (WEIGHT VOTING)

At the stage of calculating the weight voting value obtained by entering the Validity value and Euclidean value.

$$W(i) = Validitas(i) \times \frac{1}{de(1)+0{,}5}$$
$$= \frac{2}{3} x \frac{1}{de(1)+0{,}5}$$
$$= \frac{2}{3} x \frac{1}{56{,}3560+0{,}5}$$
$$= 0.011725$$

Perform the same calculation for all training data. The results of the weight voting calculation for all training data.

Table. 4. Weight voting calculation results

No	Weight Voting
1.	0,011725
2.	0,012265
3.	0,012265
4.	0,012265
5.	0,012579
6.	0,012947

No	Weight Voting
7.	0,015276
8.	0,011725
9.	1,133603
10.	0,011912
11.	0,013037
12.	1,303715
13.	0,012579

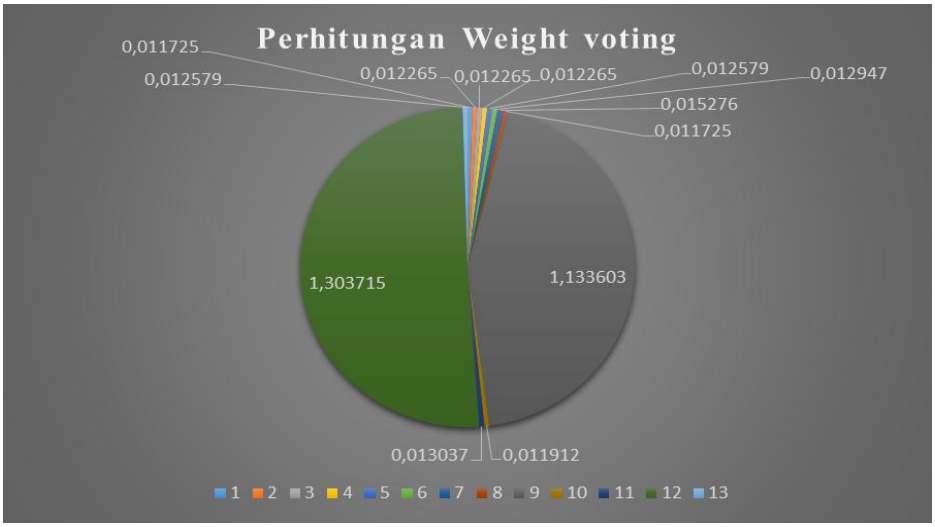

Figure 2. Diagram Pie Weight voting.

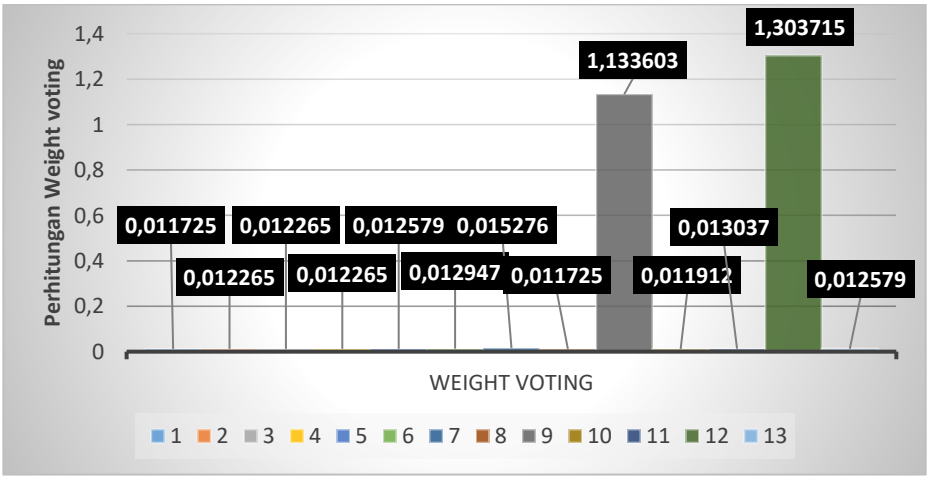

Figure 3. Weight voting Bar Chart

3.5. DISCUSSION

The results of the study showed that the measurement of the level of satisfaction of the Samsat Online Web service at Samsat Majalengka obtained a total of 447 satisfied, this shows that the results of the study on User Satisfaction of the Samsat Online website at Samsat Majalengka according to respondents are included in the high category.

However, there were still respondents who stated that they were not satisfied with a total value of 203 on the statement that the information published was accurate. In addition, respondents who stated that they were not satisfied with a total value of 58% regarding the statement that the Samsat information you need can be accessed through the Samsat Majalengka website and respondents who stated that they were satisfied with a total value of 80% regarding the statement that the Samsat Majalengka website is easy to use.

Respondents who stated that they were dissatisfied with the statement that the Samsat information you need can be accessed through the Majalengka Samsat website could be caused because according to respondents, there is sometimes a discrepancy between the information conveyed in the application and the reality in the field. Respondents who stated that they were satisfied with the statement that the Majalengka Samsat website is easy to use could be caused because according to respondents, not all application users, especially the general public, understand how to use the application, especially in the online registration menu.

CONCLUSIONS AND SUGGESTIONS

Web Samsat Online provides convenience for users to find the information they need. Proven by the results of the questionnaire given to respondents, where the level of satisfaction of Web Samsat Online users obtained a total score of 447, this shows that the results of the study on Web Samsat Online User Satisfaction at Samsat Majalengka according to respondents are in the high category. Here are suggestions: Evaluate Web Samsat Online, especially regarding data accuracy so that application users get the information they need accurately. In addition, a tutorial on using the Web is needed to help users run the application.

THANKS TO

This section is only written if desired. Contains thanks and appreciation to parties who are considered to have contributed significantly to the completion of the research.

BIBLIOGRAPHY

Dalam Pangestu dan Rusmana, (2012). Theory of Planned Behavior

Bramer, Max. 2007. Principles of Data Mining. London : Springer

Ghozali, Imam Prof Dr M.Com Akt. 2013. Aplikasi Analisis Multivariate dengan Program IBM SPSS 21. Semarang: Universitas Diponegoro.

Han, J.,&Kamber, M. 2006. Data Mining Concept and Tehniques. San Fransisco : Morgan Kauffman.

Kusrini&Luthfi,E.T. 2009. Algoritma Data Mining. Yogyakarta : Andi Publishing.

Liao. 2007. Recent Advances in Data Mining of Enterprise Data : Algorithms and Application. Singapore : World Scientific Publishing

Maimon, Oded&Rokach, Lior. 2005. Data Mining and Knowledge Discovey Handbook. New York : Springer

Noerlina. 2007. Perancangan Sistem Informasi Berbasis Object Oriented. Jakarta : Mitra Wacana Media.

Rivai, Veithzal.,&Veithzal, Andria Permata. 2006. Credit Management Handbook. Jakarta : Raja GrafindoPersada Witten, I. H., Frank, E., & Hall, M. A. 2011. Data Mining: Practical Machine Learning and Tools. Burlington :

Vercellis, Carlo. 2009. Business Intelligent: Data Mining and Optimization for Decision Making. Southern Gate, Chichester, West Sussex : John Willey& Sons, Ltd.