PAPER • OPEN ACCESS

Sentiment Analysis Review Of Smartphones With Artificial Intelligent Camera Technology Using Naive Bayes and n-gram Character Selection

View the article online for updates and enhancements.

IOP ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection−download the first chapter of every title for free.

Sentiment Analysis Review Of Smartphones With Artificial Intelligent Camera Technology Using Naive Bayes and n-gram Character Selection

R Aulianita[1*], **LA Utami**[2], **N Musyaffa**[3], **G Wijaya**[4], **A Mukhayaroh**[5], **A Yoraeni**[6]

Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri

E-mail: `rizki.rzk@nusamandiri.ac.id`

Abstract. Mobile has become a basic necessity at this time. Everyone certainly has a cellphone according to their daily needs. To capture connections and carry out various activities with just one hand. The object of this research is a review of smartphones that have the best artificial intelligent cameras. Data processing methods used in research using the Naïve Bayes algorithm. Naïve Bayes is known as one of the methods with the best classification accuracy results for text mining. The research objective is to facilitate customers who will buy a smartphone with the best AI camera without having to read product reviews. So that it can see based on the classification of positive text and label negative text classification. In this study, n-gram is used as a character selector to provide better accuracy results. Based on the results of research conducted, the accuracy of Naïve Bayes results is 72.00%, then Naïve Bayes with n-gram selection accuracy is N-gram = 2, 72.00% accuracy results, n-gram = 3, 75.00% accuracy results, and n-gram = 4 accuracy results 74.50%. In this study, carried out 10 times the experiment to measure the increased accuracy of the addition of n grams. Thus concluding that the application of the n-gram character can increase the accuracy of the Naïve Bayes algorithm.

1. Introduction

Nowadays technology is developing very fast. Some artificial intelligence (AI) technologies are widely embedded in smartphones output in 2017, such as facial recognition biometrics, augmented reality, machine learning, and so on. Apparently, AI trends in smartphones will continue to develop The popularity of cameras is growing rapidly and is being implemented for digital cameras and smartphones [1], smartphone manufacturers finally competed to make artificial intelligence camera features for their products. Until now, Chinese smartphone manufacturers are competing to create cellphones which has AI especially on Smartphone cameras, electronics and technology manufacturers such as Huawei, Honor and Xiaomi created by manufacturers in China [2]. The large selection of smartphone products that offer artificial intelligence features makes the Camera so that makes the public make a review of the product smartphone on the market. Internet can provides a variety of text data that can be used become useful information to customers [3]. Sentiment analysis has become the focus of research for many researchers who take the field text mining research and focus to natural language processing and social network analysis [4.] The large effects and benefits of sentiment analysis

cause research-based application and sentiment analysis to grow rapidly [5]. The methods most widely used in sentiment analysis are Naïve Bayes (NB), Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) [6]. Naïve Bayes Classifier is the best method for the text classification algorithm to improve performance of the classification.

The advantage of Naïve Bayes is the Naif-Bayes classification assumes that variable are independent. For example mobile features, camera, mobile, simcard, wifi. The Naïve Bayes Clasifier assumes that there are no mobile features related to each other. It is assumed that there is no relationship between a particular event from a group with the presence or absence of other events. So Naïve Bayes has the best performance [7]. Naive Bayes is an approach to classify models based on assigned labels, the simplest and most successful technique for large-scale datasets [8].

Classification is done for two variables called aspects and sentiments. Explaining the system in the generative model, we can say that both variables affect the use of words in sentences. Therefore, the probability distribution of words depends on the value of each variable [9], the N-gram probabilistic model, which is a model used to predict the next possible word from the previous N-1 word [10] n-gram is a slice of the number of n characters in a string. N-gram is a method that is applied for word or character generation [11].

In this study, it was determined that data processing was performed using the Naïve Bayes Algorithm. The results of the Naïve Bayes accuracy are compared with n-gram characters. The hypothesis in this study is whether the character n-gram can affect the accuracy of the Naïve Bayes method. This study is to measure the increase accuracy based on n gram charter selections using Naïve Bayes method.

2. Related Work

2.1. Sentiment Analysis
Sentiment analysis or can be called mining opinion representing public studies of services, products, film services and so forth. This sentiment analysis can be used to consider decision making so that accurate decisions are made. Sentiment analysis is widely used for the benefit of text mining research. In a business industry, text mining is used to view reviews or opinions of each customer, which is useful for providing information for customers. This information is useful as a determinant in making business decisions [12].

The development of various social media platforms such as Facebook, Twitter, Instagram, allows anyone to write opinions or comments so quickly in a short time. So they can express opinions through text or emoticons. Text mining, is very important in taking the text to be processed into a right decision for business people [13].
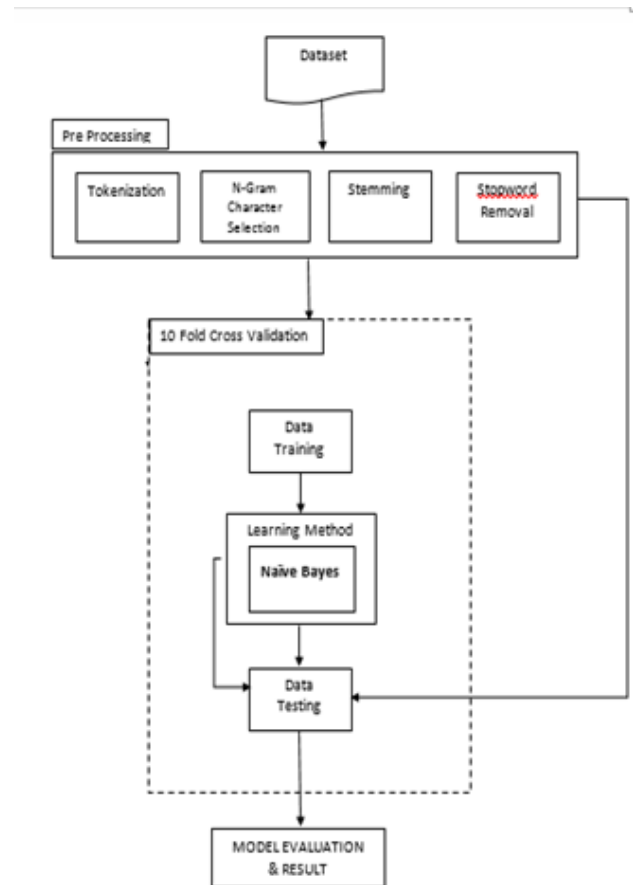
2.2. Naïve Bayes
Naïve Bayes algorithm is used to take the steps to process the data to obtain the result of sentiment analysis. Naïve Bayes algorithm is a classification technique most commonly used for mining opinion [14]. When processing training data, Naïve Bayes produces accuracy, precision and recall, the results of that accuracy are used to analyze the sentiment of each review. Based on the results of previous studies, it is proven that Naïve Bayes works better, which is shown through the results of accuracy [15]. Naive-Bayes classification assumes that its features are not related to each other in general. To help identify objects, features can be defined as attributes.

2.3. N-gram Character Selection
N-gram is a set of words given in a paragraph and when counting n-grams is usually done by moving one word forward [16]. Additional preprocessing steps for the bag-of-words approach included the n-gram technique and term frequency-inverse document frequency vectorization [17]. in that study provides an increase in accuracy for the use of n grams [18].

3. Research Method

3.1. Propose Model

Figure 1. Propose Model Naïve Bayes with n-gram character selection.

The dataset obtained was 200 datasets. Each of the 100 positive AI camera reviews and 100 negative AI cameras. Based on the existing dataset, then pre-processing is done using Tokenization, N-Gram, Stemming and Stopword Removal. The validation used in this study is Cross Validation, exactly, 10 times the change in n grams to evaluate perform of accuracy n gram. Learning method used is Naïve Bayes, then the training data is done testing and ends on the result or Model Data Evaluation.

3.2. Research Methodology

(i) Data Collecting
   The data used in this study is a set of smartphone product reviews with artificial intelligence camera technology collected from amazon.com with the keywords "Realme Camera AI", "Huawei Camera AI", and "Xiaomi Camera AI". Since March 2020, we have collected 200 smartphone product reviews under the Realme, Huawei and Xiaomi brands. Online reviews were posted by more than 2000 reviewers (customers) of 3 brands of mobile phones. Each review taken consists of the following information: 1) ID of the reviewer; 2) product ID; 3) ranking; 4) time of review; 5) location of reviewers; 6) review the text. After data

collection, all reviews are labeled manually. 200 reviews will be classified into two categories, 100 positive reviews and 100 negative reviews according to the three smartphone brands. The format used in the collection of .txt. We have converted the training and test data into English transliterated before we use to data training and testing.

(ii) Data Preprocessing
Before the dataset is ready for use, the data is preprocessing first so that the dataset is clean and ready to be used in the next process. The preprocessing data used for this dataset are:

(a) Tokenization
The process of cutting every word in the text and changing the letters in the document into lowercase letters. Only letters are accepted, special characters or punctuation will be removed.

(b) N-Gram Character n-gram is assumed to improve the performance of a Naïve Bayes algorithm. Based on the results of previous studies, n-gram can increase the accuracy value in the Naïve Bayes method. N-gram is a set of words given in a paragraph and when counting n-grams is usually done by moving one word forward (although in the process there is a process in which words are advanced a number of Y words).

(c) Stemming Search for the basic words of a word by removing the affix contained in the word.

(d) Stopwords Removal The process of eliminating words that often appear but do not have any influence in the extraction of sentiments of a review such as a timepiece, or question word.

(iii) Classification
The classification process is done by using an application and the method used is Naïve Bayes Classifier. There are two stages carried out in this process, the first stage is the training process (training) and the second stage is the testing process (testing).

(a) Training
The training phase will process the formation of classes and as a reference for how documents will be classified.

(b) Testing
This process is used to determine the accuracy of the model built in the training process, commonly used data called a test set to predict labels.

(iv) Evaluation
The process performance of the Naïve Bayes method will be assessed using k-fold cross validation, while the accuracy of the algorithm will be represented using the Confusion Matrix.

4. Experiment and Result

Sentiments that represent positive opinions in this study such as Satisfied, Good, Excellent, while to represent negative reviews are bad, not and no recommend. Validation used 10-fold cross validation for the Naïve Bayes design model with n-gram. The results of testing the data by the Naïve Bayes method are:

Table 1. Result Experiment Of Naïve Bayes.

|	NB	AUC
Accuracy	72.00%	0.728.

Based on data processing carried out that the results of the accuracy of Naïve Bayes amounted to 72.00%.

In accordance with the hypothesis of researchers that by applying the character n-gram into the Naïve Bayes algorithm that can improve the performance of the method. It was proven when given n-gram = 2 that the resulting accuracy was 72.00%, then n-gram was added to n-gram = 3. The accuracy was 75.00% and n-gram = 4 showed an accuracy of 74.50%. As for the difference, it can be seen in the table below.

Table 2. NB Accuracy Value with n gram evaluation.

n-gram	Accuracy	AUC
1	72.00%	0.728.
2	72.00%	0.728.
3	75.00%	0.750.
4	74.50%	0.745.
5	70.50%	0.705.
6	75.00%	0.750.
7	74.00%	0.740.
8	73.00%	0.730.
9	73.50%	0.735.
10	75.50%	0.755.

Based on the above experiments, it was concluded that the n-gram character has the ability to be able to improve the accuracy of Naïve Bayes results, after testing it was proven that it can improve the accuracy and AUC results when adding n grams.
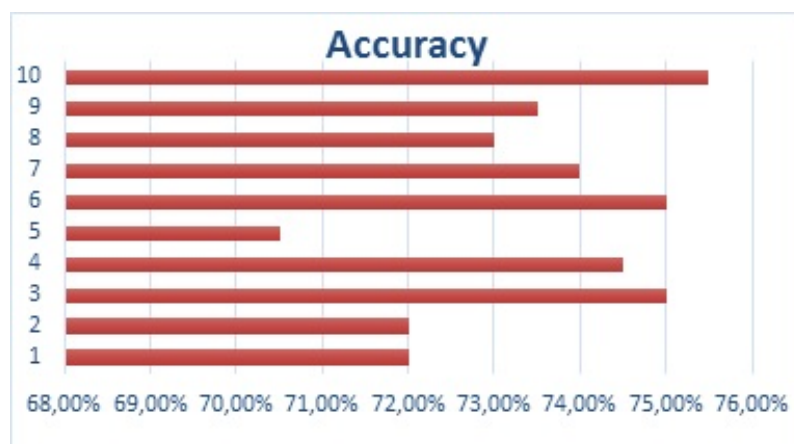
Figure 2. Accuracy NB + n gram evaluation.

5. Conclussion and Discussion

The results of testing with the Naïve Bayes model show that this method is one of the simplest methods used for text classification. This test uses a positive AI camera review and a negative

AI camera review. The accuracy of the test using the Naïve Bayes algorithm was 72.00%. The application of n-gram is proven to be able to improve the accuracy results by applying n-gram = 2 producing 72.00% accuracy, n-gram = 3 producing 75.00% and n-gram = 4 producing 74.50%. Based on the above data processing that n grams can affect the level of accuracy. This was proven after testing data training 10 times on a sentiment review camera with artificial intelligent technology able to increase the value of accuracy.

References

[1] Takemura Y, 2019 The Development of Video-Camera Technologies: Many Innovations behind Video Cameras Are Used for Digital Cameras and Smartphones IEEE Consum. Electron. Mag. 8, 4 p. 10–16.

[2] Wardani A, 2017, 5 Negara Asal Vendor Smartphone Terkemuka Dunia - Tekno Liputan6.com.

[3] Ernawati S Yulia E R Frieyadie and Samudi, 2019 Implementation of the Naïve Bayes Algorithm with Feature Selection using Genetic Algorithm for Sentiment Review Analysis of Fashion Online Companies 2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018 Citsm p. 1–5.

[4] Anitha A and Sc A M, 2019 A COMPARATIVE STUDY AND SENTIMENT ANALYSIS USING KONSTANZ INFORMATION 6 p. 488–494.

[5] Buntoro G A, 2017 Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter Integer J. Maret 1, 1 p. 32–41.

[6] Kristiyanti D A Umam A H Wahyudi M Amin R and Marlinda L, 2019 Comparison of SVM Naïve Bayes Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter 2018 6th Int. Conf. Cyber IT Serv. Manag. CITSM 2018 Citsm p. 1–6.

[7] Ibrahim M N and Yusoff M Z, 2017 IEEE Conference on Open Systems (ICOS): 13th-14th November 2017, Miri Marriott Resort and SPA, Miri, Sarawak, Malaysia Impact Differ. Train. Data Set Accuarcy Sentim. Classif. Naive Bayes Tech. November 1 p. 17–20.

[8] Jadon P Bhatia D and Mishra D K, 2019 A BigData approach for sentiment analysis of twitter data using Naive Bayes and SVM Algorithm IFIP Int. Conf. Wirel. Opt. Commun. Networks, WOCN 2019-Decem.

[9] Mubarok M S Adiwijaya A and Aldhi M D, 2017 Aspect-based sentiment analysis to review products using Naïve Bayes AIP Conf. Proc. 1867, August.

[10] Suhartono D, 2019, N-Gram.

[11] Chandra D N Indrawan G and Sukajaya I N, 2019 Klasifikasi Berita Lokal Radar Malang Menggunakan 2.

[12] Marrese-Taylor E Velásquez J D Bravo-Marquez F and Matsuo Y, 2013 Identifying customer preferences about tourism products using an aspect-based opinion mining approach Procedia Comput. Sci. 22 p. 182–191.

[13] S. S and K.V. P, 2020 Sentiment analysis of malayalam tweets using machine learning techniques ICT Express xxxx p. 2–7.

[14] Annett M and Kondrak G, 2008 A Comparison of Sentiment Analysis Techniques: Polarizing Movie Blogs Figure 1 p. 25–26.

[15] Fitri V A Andreswari R Hasibuan M A Fitri V A Andreswari R and Hasibuan M A, 2019 ScienceDirect ScienceDirect ScienceDirect Sentiment Analysis of Social Media Twitter with Case of Anti- Sentiment Analysis of Social Media Twitter with Case of Anti- LGBT Campaign in Indonesia using Naïve Bayes , Decision Tree , LGBT Campaign in Indonesia Procedia Comput. Sci. 161 p. 765–772.

[16] Sarkar K, 2018 Using Character N-gram Features and Multinomial Naïve Bayes for Sentiment Polarity Detection in Bengali Tweets Proc. 5th Int. Conf. Emerg. Appl. Inf. Technol. EAIT 2018 p. 1–4.

[17] Nguyen V H Nguyen H T Duong H N and Snasel V, 2016 n -Gram-Based Text Compression 2016.

[18] Indrayuni E and Wahyudi M, 2015 PENERAPAN CHARACTER N-GRAM UNTUK SENTIMENT ANALYSIS REVIEW HOTEL MENGGUNAKAN ALGORITMA NAIVE BAYES 8.