SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 3, Number 2, April 2019
DOI : https://doi.org/10.33395/sinkron.v3i2.242
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Naïve Bayes Algorithm For Sentiment Analysis Windows Phone Store Application Reviews

Normah
STMIK Nusa Mandiri Jakarta: Teknik Informatika
Jakarta, Indonesia
normah.nor@nusamandiri.ac.id

***Abstract*—** Reading reviews helps consumers choose the applications, helping companies and developers monitor user satisfaction to improve quality of features and services, read overall and manually could spend the time and laborious, if read at a glance, information not conveyed perfectly. This study analyzes user sentiment Windows Phone Store applications by automatically classifying reviews into positive or negative opinion category. Naïve Bayes has good potential because of its simplicity and performance as a model of classifying text on many domains. The model was evaluated using 10 Fold Cross Validation. Measurements were made with the Confusion Matrix and the ROC curve. The accuracy produced in this study is 84.50%, indicating that Naïve Bayes is a good model in classifying text especially in the case of sentiment analysis.

Keywords—Naïve Bayes; Sentiment Analisyst; Windows Phone Store; Application Review.

I. INTRODUCTIVON

Windows Phone Store (www.windowsphone.com) is not different from other e-commerce sites, in addition to providing some free content application, also offered a paid application to run on smartphones based on the Windows Phone. In the world of market applications, Developer commonly provide specifications and description of the features of the application they offers – interoperable with application content display (screen shots), product reviews, rating or star ratings from users based on their experience after installing and using the application.

The contents of the reviews was very influential to reputation of the application, as well as the the developer's reputation as a seller or manufacturer of the application, because only by reading it we can determine that the application is good or not from the user side. From the reviews, the consumers can find out how the quality of an application, helping consumers in determining and selecting which application best from dozens of similar available applications, helping the developer to monitor user satisfaction, as an input in improving quality (update) the features and services of the application they had made. Read the whole review manually can take time and effort, but if you read just a glance, most feared, information is not delivered well. To overcome it, needed a research to analyze the user sentiment against Windows Phone application Store through the reviews from the users of the Windows Phone application Store by automatically classify these reviews into two categories, namely a positive or negative opinion.

However, classify documents into specific categories properly and with a high degree of accuracy still become a challenge, because the number of features in the dataset is very large and spacious. Naïve Bayes have a good potentially and widely used as a model for the classification of the document because of its simplicity and have a good performance in the stage of training and at the stage of classification (Ting, Ip, & Tsang, 2011). In this research was conducted to classification to analyse the text in which the sentiments against the reviews written by users of Windows Phone products application Store using the algorithm approach Naïve Bayes.

In the research journal “Understanding Online Consumer Review Opinions with Sentiment Analysis using Machine Learning” (Yang, Tang, Wong, & Wei, 2010), say that with the advent of Web 2.0 technologies, the Web has evolved to become a popular channel of communication and interaction between Web users and online consumers. Social media, unlike traditional media, have rich but unorganized content contributed by users, often in fragmented and sparse fashion. Users usually spend a lot of their time filtering useless information and yet are not able to capture the essence. In this study, we focus on user-contributed reviews of products, which many online consumers use to support their purchase

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 3, Number 2, April 2019
DOI : https://doi.org/10.33395/sinkron.v3i2.242
e-ISSN : 2541-2019
p-ISSN : 2541-044X

decisions by identifying products that best fit their preferences. In the recent years, sentiment classification and analysis of online consumer reviews has drawn significant research attention. Most existing techniques rely on natural language processing tools to parse and analyze sentences in a review, yet they offer poor accuracy, because the writing in online reviews tends to be less formal than writing in news or journal articles. Many opinion sentences contain grammatical errors and unknown terms that do not exist in dictionaries. Therefore, this study proposes two supervised learning techniques (class association rules and naïve Bayes classifier) to classify opinion sentences into appropriate product feature classes and produce a summary of consumer reviews. An empirical evaluation that compares the performance of the class association rules technique and the naïve Bayes classifier for sentiment analysis shows that our proposed techniques achieve more than 70% of the macro and micro F-measures.

In the research journal "Do Android Users Write About Electric Sheep?" (Ha & Wagner, 2013), say that Consumer reviews and star ratings are integral to application markets. The content of reviews help consumers determine whether an application is "good" or not. Since consumers rely heavily on reviews when selecting applications, we wanted to know what was being written about in reviews. In particular, we wanted to know if users were discussing privacy and security risks of an application, and if not, what were they writing about instead? In our work, we manually analyzed Android users' reviews to see what they write about when reviewing Google Play applications. Overall, only 1% of our reviews mentioned application permissions. We also found that a small subset of reviews relating to preinstalled applications and applications that requested a user's rating had underlying privacy and security implications. The majority of reviews focused on the quality of applications: people often described an application using an adjective (e.g., "great app" or "horrible"), wrote about its feature/functionality, specifically said if the application worked or not, and/or put their phone or tablet model in the review. We also found that sentiment did influence reviewers' ratings of the applications. In general, the overall star rating of our sample was overwhelmingly positive, suggesting that Google Play is no different from other e-commerce sites.

II. LITERATURE REVIEW

A. *Text Mining*

Text mining is one special field data mining (Feldman, 2013). Text mining is the process of extracting patterns in the form of information and knowledge that is useful from a large number of text data source that aims to find the words that can represent the contents of the document so that it can be done analysis of the connectedness between documents. The usual stages done as follows (Harlian, 2006):

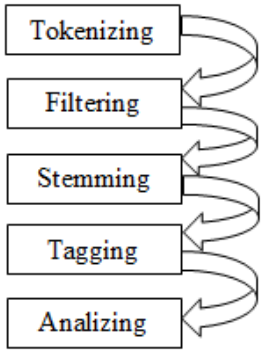

Figure 1. Stages of Text Mining

1. Tokenizing
   Collect all the words that appear in the document, then eliminate any punctuation marks and symbols that are not letters.
2. Filtering
   Retrieve and stored important words from the word list tokens using the algorithm, and discard the irrelevant words or less essential to use algorithms stop list (stopword removal).
3. Stemming
   The process of classifying words to be formationing derived from the same root.
4. Tagging
   A Phase to seek early form / root of each word or word past results stemming..
5. Analyzing (Weight Calculation word)
   The process of calculating the weight (w) of documents in order to know how far the degree of similarity between the keywords entered by the document

B. *Sentiment Analysis*

Sentiment analysis is the extraction of information from text data sources to detect positive or negative views of an object. Usually applied to identify the trend of public opinion on a product or company. The Steps to text classification sentiment analysis are (Moraes, Valiati, & Neto, 2013):

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 3, Number 2, April 2019
DOI : https://doi.org/10.33395/sinkron.v3i2.242
e-ISSN : 2541-2019
p-ISSN : 2541-044X

1. Define domain datasets
   The collection of datasets in a spanning domain.
2. Pre-processing
   Initial processing steps are generally performed tokenization, Transform case, stopwords removal filter, Filter Tokens, and stemming.
3. Transformation
   The process of representation numbers calculated from textual data. Binary representation is commonly used and only count the presence or absence of a word in the document. How many times a word appears in a document is also used as a weighting scheme of textual data.
4. Feature Selection
   Reducing the amount of data to be analyzed by identifying relevant features.
5. Classification
   The process of classifying data into certain categories
6. Validation
   The Evaluating process accuracy of the result prediction model
7. Interpretation/Evaluation
   Phase counting accuracy, recall, precision, and F-1.

C. *Naïve Bayes algorithm*

Naïve Bayes is a learning algorithm that is often used to overcome the problem of text classification, is one of the machine learning method that uses probability calculations. This method utilizes the theory put forward by the British scientist Thomas Bayes 8, which predicts the probability in the future based on past experience. Bayes Theorem is a theorem which is used in statistics to calculate the odds for a hypothesis, Bayes Optimal Classifier calculates the probability of a class of each group attributes exist, and determine which one is the most optimal class:

D. *Evaluation and Validation Algorithm*

Validation is the process of evaluating the accuracy of the results of the prediction model. K-Fold Cross Validation is a validation technique that divides the data into k parts and then each part will be the classification process. By using K-Fold Cross Validation experiments will be carried out as many as k. Each experiment will use one data testing and K-1 part will become a training data, then the testing data will be exchanged with one of training data, so for each experiment will be get a different data testing. To test the model, use the method Confusion Matrix, and the ROC curve.

1. Confusion Matrix
   used to analyze how well the classifier can identify tuples of different classes.
2. ROC Curve
   ROC curve (Receiver Operating Characteristic) shows the classification accuracy and compare visually. ROC express confusion matrix. ROC is a two-dimensional graph with false positives as horizontal lines and true positives to measure the performance difference method is used. ROC curves are used to measure the AUC (Area Under the Curve). AUC was calculated to measure the difference performances method was used. ROC curve divides the positive results in the y-axis and the negative results in the x-axis (Witten & Frank, 2011). So the larger the area under the curve, the better the prediction results. AUC values were divided into several groups (Gorunescu, 2011):
   - 0.90 - 1.00 = Excellent Classification
   - 0.80 - 0.90 = Good Classification
   - 0.70 - 0.80 = Fair Classification
   - 0.60 - 0.70 = Poor Classification
   - 0.50 - 0.60 = Failure

III. RESEARCH METHODOLOGY

The method which used is the method of experimental research with research phases as follows: Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

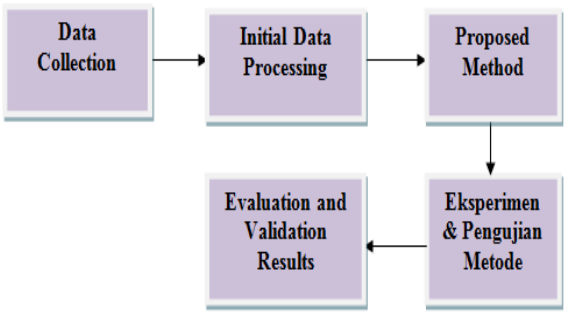

Figure 2. Stages of Research

In this study sentiment analysis of product reviews of Windows Phone Store applications using the Naïve Bayes algorithm approach. Review of data has been collected through a preprocessing stage in advance to obtain relevant words to be classified. The evaluation process is carried out using 10-fold cross validation. Measurement accuracy is measured by the confusion matrix. The following steps are carried out research:

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 3, Number 2, April 2019
DOI : https://doi.org/10.33395/sinkron.v3i2.242
e-ISSN : 2541-2019
p-ISSN : 2541-044X

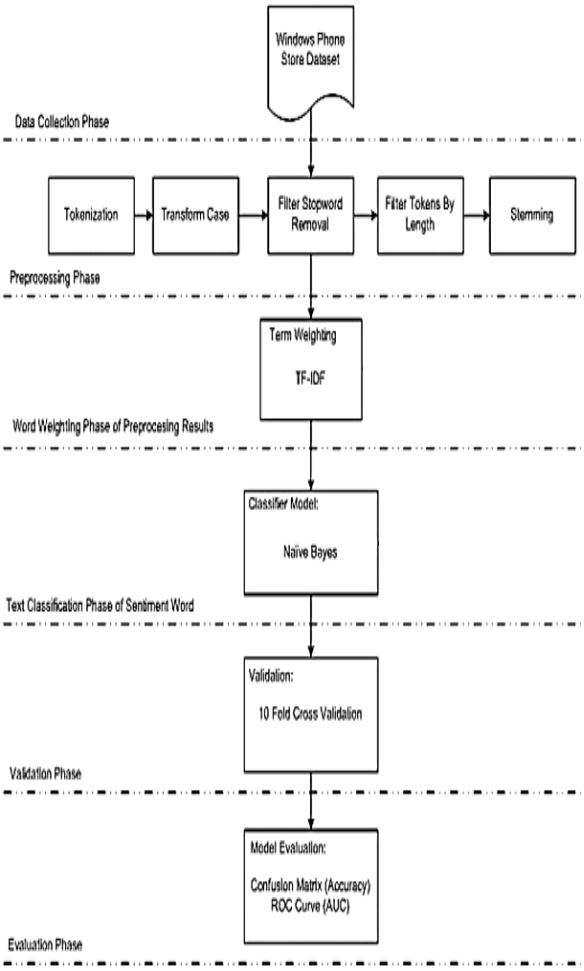

Figure 3. Steps of Research

IV. RESULTS AND DISCUSSION

The method which used is the method of experimental research with research phases as follows: Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. Data Collection Phase

The Research data in the form of 150 positive reviews and 150 negative reviews in English from the users of the product application Windows Phone Store, Where 100 positive reviews and 100 negative reviews will be used at the stage of training and data testing on classification model, as many as 50 positive reviews and 50 negative reviews will be tested in applications designed for implementation. Data reviews were collected will be in the separated storage per review in the form of document notepad with .txt as the extension.

B. Initial Data Processing Phase

Stages of the preprocessing performed i.e.:

1. Tokenization
   All the words in each document reviews are collected and punctuation marks, and symbols were removed.

TABLE I. COMPARISON TEXT BEFORE AND AFTER TOKENIZATION

Before Tokenization	After Tokenization
Terrible controls, which basically make the game unplayable. Game is gliched, so it's difficult to play through some levels, with the mission constantly changing on you. It's unclear half the time where you should go and what you need to do. Not a lot of options when fighting, so you just tap the same button as quickly as possible...this gets really boring. Basically this is a frustrating and dull game. Worst of all, I paid for this to find that there are micro-transactions!!! Are you kidding me?	Terrible controls which basically make the game unplayable Game is gliched so it s difficult to play through some levels with the mission constantly changing on you It s unclear half the time where you should go and what you need to do Not a lot of options when fighting so you just tap the same button as quickly as possible this gets really boring Basically this is a frustrating and dull game Worst of all I paid for this to find that there are micro transactions Are you kidding me

2. Transform Case
   All uppercase in the document reviews converted into lower case.

TABLE II. COMPARISON TEXT BEFORE AND AFTER TRANSFORM CASE

Before Transform Case	After Transform Case
Terrible controls which basically make the game unplayable Game is gliched so it s difficult to play through some levels with the mission constantly changing on you It s unclear half the time where you should go and what you need to do Not a lot of options when fighting so you just tap the same button as quickly as possible this gets really boring Basically this is a frustrating and dull game Worst of all I paid for this to find that there are micro transactions Are you kidding me	terrible controls which basically make the game unplayable game is gliched so it s difficult to play through some levels with the mission constantly changing on you it s unclear half the time where you should go and what you need to do not a lot of options when fighting so you just tap the same button as quickly as possible this gets really boring basically this is a frustrating and dull game worst of all i paid for this to find that there are micro transactions are you kidding me

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 3, Number 2, April 2019
DOI : https://doi.org/10.33395/sinkron.v3i2.242
e-ISSN : 2541-2019
p-ISSN : 2541-044X

3. Filter Stopwords Removal
   Unrelevant words in the document reviews will be deleted, for example, the, of, for, with, and other words that have no meaning if separated by other words and not associated with adjectives related to sentiment word.

TABLE III. COMPARISON TEXT BEFORE AND AFTER STOPWORD REMOVAL

Before Stopword Removal	After Stopword Removal
terrible controls **which** basically make **the** game unplayable game **is** gliched **so it** s difficult **to** play **through some** levels **with the** mission constantly changing **on you it** s unclear half **the** time **where you should** go **and what you need to do not a** lot **of** options **when** fighting **so you just** tap **the same** button **as** quickly **as possible this** gets **really** boring basically **this is a** frustrating **and** dull game worst **of all** i paid **for this to** find **that there are** micro transactions **are you** kidding **me**	terrible controls basically make game unplayable game gliched s difficult play levels mission constantly changing s unclear half time go lot options fighting tap button quickly gets boring basically frustrating dull game worst i paid find micro transactions kidding

4. Filter Token By Length

   Delete the word in the document review did not reach the minimum limit character and exceed the maximum limit character. On this research, the minimum limit used 3 characters and maximum limit used 25 characters.

TABLE IV. COMPARISON TEXT BEFORE AND AFTER FILTER TOKENS BY LENGTH

Before Filter Tokens By Length	After Filter Tokens By Length
terrible controls basically make game unplayable game gliched **s** difficult play levels mission constantly changing **s** unclear half time **go** lot options fighting tap button quickly gets boring basically frustrating dull game worst **i** paid find micro transactions kidding	terrible controls basically make game unplayable game gliched difficult play levels mission constantly changing unclear half time lot options fighting tap button quickly gets boring basically frustrating dull game worst paid find micro transactions kidding

5. Stemming

   Looking for the roots of each of the words in the document reviews, cut out each word variants to find base words and classifies words that derive from the same base, such as write, wrote and written where the base word is write.

TABLE V. COMPARISON TEXT BEFORE AND AFTER STEMMING

Before Stemming	After Stemming
terrible controls basically make game **unplayable** game **gliched** difficult play **levels** mission **constantly changing** unclear half time lot **options fighting** tap button **quickly gets boring basically frustrating** dull game worst paid find micro **transactions kidding**	terribl control basic make game unplay game glich difficult play level mission constant chang unclear half time lot option fight tap button quick get bore basic frustrat dull game worst paid find micro transact kid

C. *Initial Data Processing Phase*

At this stage done the calculation of the weighting of words resulting from the preprocessing algorithm using TF-IDF.

ExampleSet (200 examples, 4 special attributes, 1203 regular attributes)

Row No.	label	metadat...	metadata_p...	metadata_d...	aap	abil	abl	absolut	abysm
1	negatif	doc1.txt	D:\tesis_win	14 Jul 14 7:4	0	0	0	0	0
101	positif	doc1.txt	D:\tesis_win	21 Agu 14 1:	0	0	0	0	0
2	negatif	doc10.txt	D:\tesis_win	14 Jul 14 9:5	0	0	0	0	0
102	positif	doc10.txt	D:\tesis_win	14 Jul 14 9:3	0	0	0	0	0
3	negatif	doc100.txt	D:\tesis_win	21 Agu 14 1	0	0	0	0	0
103	positif	doc100.txt	D:\tesis_win	19 Agu 14 1	0	0	0	0	0
4	negatif	doc11.txt	D:\tesis_win	14 Jul 14 10	0	0	0	0	0
104	positif	doc11.txt	D:\tesis_win	21 Agu 14 1:	0	0	0	0	0
5	negatif	doc12.txt	D:\tesis_win	14 Jul 14 10	0	0	0	0	0
105	positif	doc12.txt	D:\tesis_win	14 Jul 14 9:5	0	0	0	0	0
6	negatif	doc13.txt	D:\tesis_win	14 Jul 14 10	0	0	0	0	0
106	positif	doc13.txt	D:\tesis_win	14 Jul 14 9:5	0	0	0.232	0	0
7	negatif	doc14.txt	D:\tesis_win	14 Jul 14 10	0.359	0	0	0	0
107	positif	doc14.txt	D:\tesis_win	14 Jul 14 10	0	0	0	0	0
8	negatif	doc15.txt	D:\tesis_win	14 Jul 14 14	0	0	0	0	0

Figure 4. Results of The Weighting Words Using TF-IDF Algorithm

D. *Sentiment Text Classification Phase Using Naïve Bayes Algorithm*

Text classification is performed to determine whether a review is included as a positive class or negative class based on the greater calculation of the

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 3, Number 2, April 2019
DOI : https://doi.org/10.33395/sinkron.v3i2.242
e-ISSN : 2541-2019
p-ISSN : 2541-044X

probability of the Naïve Bayes Algorithm formula. The presence of words in a document reviews will be represented by the numbers 1 and 0 if the word did not appear in document reviews.

TABLE VI. VECTOR DOCUMENTS BOOLEAN AND LABEL CLASS CLASSIFICATION RESULTS

Doc	1	20	30	44	87	49	69	71	85	93
Annoy	0	0	0	0	0	0	1	0	1	0
bore	0	0	0	0	0	1	0	1	0	0
Disappoint	0	0	0	0	0	0	1	0	0	1
Terribl	0	0	0	0	0	1	0	0	0	1
Worst	0	0	0	0	0	1	0	1	1	0
Awesome	0	0	0	1	1	0	0	0	0	0
Fun	1	0	0	0	1	0	0	0	0	0
Good	1	0	1	0	0	0	0	0	0	0
Great	1	1	1	0	1	0	0	0	0	0
Love	0	1	0	1	0	0	0	0	0	0
Class Result	Positive					Negative				

Based on the probability of the above, it can be concluded that the document reviews doc1. txt included in class positive, because P (positive | doc1) greater than P (negative | doc1). A design model from the above calculation using RapidMiner 5.2 is:

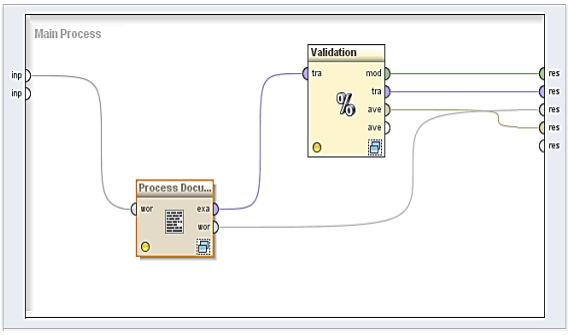

Figure 5. Naïve Bayes Classification Model Design Using RapidMiner 5.2

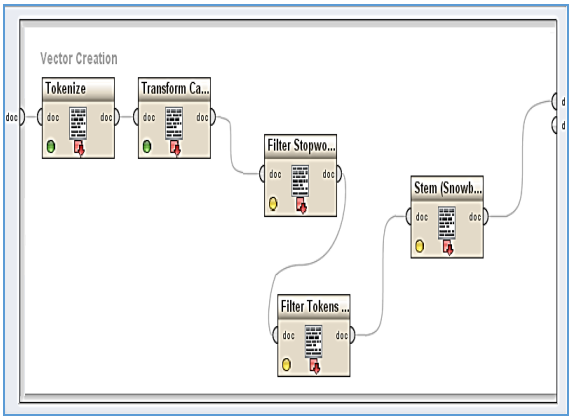

Figure 6. Preprocessing Phase

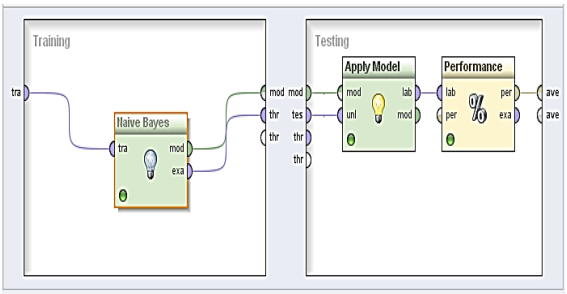

Figure 7. Validation Phase Using 10 Fold Cross Validation

E. Model Validation/Testing Phase Using 10 Fold Cross Validation

To do the testing model used techniques 10 Cross Validation, where data divided randomly into 10 parts. The testing process begins with the formation of a model with the data in the first part. The formed model will be tested on 9 formed part of the remaining data, then the process accuracy is calculated by looking at how much data has been classified correctly. Figure 6 and Figure 8 above is a picture of design models of model testing phase used technique 10 Fold Cross Validation.

F. Evaluation Phase (Measurement of the Confusion Matrix and ROC Curve/AUC)

The result of testing model will be discussed through the confusion the matrix to show how good model that is formed.

TABLE VII. CONFUSION MATRIX MODEL NAIVE BAYES METHOD

Naïve Bayes Accuracy: 84.50% +/- 5.22% (micro: 84.50%)			
	True Positive	True Negative	Class precision
Positive Prediction	90	21	81.08 %

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Journal Publications & Informatics Engineering Research
Volume 3, Number 2, April 2019
DOI : https://doi.org/10.33395/sinkron.v3i2.242
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Naïve Bayes Accuracy: 84.50% +/- 5.22% (micro: 84.50%)			
Negative Prediction	10	79	88.76 %
Class Recall	90.00 %	79.00 %	

Below is a table test results the model Algorithms Naïve Bayes:

TABLE VIII. TESTING OF ALGORITHMS NAÏVE BAYES MODEL

Successful Negative Review Classification	Successful positive Classification	Model Accuracy	AUC
90	79	84.50 %	0.866

Based on the table above can be seen that the test using Naïve Bayes classification model yield good classification, with a value of accuracy is 84.50%, and the AUC value is 0.866.

Confusion Matrix above calculation results visualized through the ROC curve.

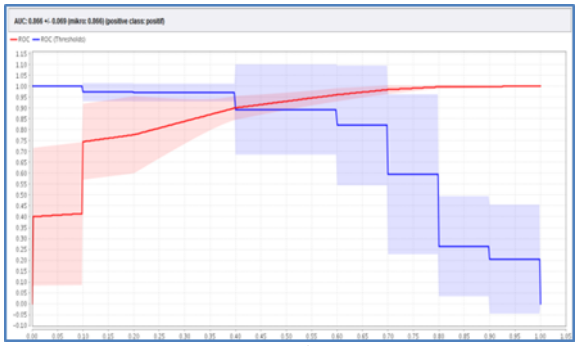

Figure 8. ROC Curve of Naïve Bayes Model

V. CONCLUSION

1. Based on the results of the model tests performed, an accuracy of 84.50% was obtained, and can be said that Naïve Bayes is indeed a good method in classifying text especially in the case of sentiment analysis as in this study.
2. The results of this study can help the companies, developers, and users applications in analyzing sentiment Windows Phone Store product applications reviews by automatically classifying reviews in English into two categories: positive and negative automatically with a short time.

REFERENCES

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, *56*(4), 82–89.

Gorunescu, F. (2011). *Data Mining: Concepts and Techniques*. Verlag berlin Heidelberg: Springer.

Ha, E., & Wagner, D. (2013). Do Android users write about electric sheep? Examining consumer reviews in Google Play. *2013 IEEE 10th Consumer Communications and Networking Conference, CCNC 2013*, 149–157. https://doi.org/10.1109/CCNC.2013.6488439

Harlian, M. (2006). Text Mining.

Moraes, R., Valiati, J. F., & Neto, W. P. G. (2013). Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Applications*, *40*(2), 621–633. Retrieved from linkinghub.elsevier.com/retrieve/pii/S0957417412009153

Ting, S. L., Ip, W. H., & Tsang, A. H. C. (2011). Is Naïve bayes a good classifier for document classification? *International Journal of Software Engineering and Its Applications*, *5*(3), 37–46.

Witten, I. H., & Frank, E. (2011). *Data Mining Practical Machine Learning Tools And Technique*. Burlington: Elsevier Inc. https://doi.org/0120884070, 9780120884070

Yang, C. C., Tang, X., Wong, Y. C., & Wei, C.-P. (2010). (2010). Understanding Online Consumer Review Opinions with Sentiment Analysis using Machine Learning. *Journal of the Association for Information Systems*, *2*(3), 73–89. Retrieved from http://aisel.aisnet.org/pajais/vol2/iss3/7/