CITSM 2018

IEEE
The Indonesia Section
Computer Society Chapter

CITSM 2018

The 6th International Conference on Information of Technology

CYBER AND IT SERVICE MANAGEMENT

INNA PARAPAT (Lake Toba)

August 7, 2018 - August 9, 2018

CONFERENCE PROGRAM

Organized By :

Partners & Co-Organizers :

2018 6th International Conference on Cyber and IT Service Management

Convention Hall, Inna Parapat Hotel

August 7-9, 2018

ISBN : 978-1-5386-5433-0

IEEE Catalog Number : CFP1837Z-PRT

2018 6th International Conference On Cyber And IT Service Management (CITSM)

Convention Hall, Inna Parapat Hotel

Phone: (0625) 41012

Email : contact.citsm@uinjkt.ac.id

Website : http://citsm.id/

August 7-9, 2018

ISBN : 978-1-5386-5433-0

IEEE Catalog Number : CFP1837Z-PRT

2018 6th International Conference On Cyber And IT Service Management (CITSM)

ISBN : 978-1-5386-5433-0

IEEE Catalog Number : CFP1837Z-PRT

Additional copies of this publication are available from

Curran Associates, Inc.

57 Morehouse Lane

Red Hook, NY 12571 USA

+1 845 758 0400

+1 845 758 2633 (FAX)

email: curran@proceedings.com

PREFACE

Assalaamu ‘alaykum warahmatullahi wabarakaatuh,

The CITSM 2018 is in the general area of communication and information technology. It provides a forum for presenting and discussing the latest innovations, results and developments in IT Management & organizations, IT Applications, Cyber & IT Security, and ICT. The main objective of this conference is to provide a forum for engineers, academia, scientist, industry, and researchers to present the result of their research activities in the field of Computer and Information Technology. The primary focus of the conference is to create an effective medium for institutions and industries to share ideas, innovations, and problem solving techniques.

There are 282 papers submission and only 150 papers are accepted and 147 papers have been registered and presented. Accepted papers will be presented in one of the regular sessions and will be published in the conference proceedings volume. All accepted papers are submitted to IEEEXplore. IEEE Conference Number: # 43622. Catalog Number: CFP1837Z-PRT, ISBN: 978-1-5386-5433-0, CFP1837Z-USB, ISBN: 978-1-5386-5434-7.

On behalf of the CITSM organizers, we wish to extend our warm welcome and would like to thank for the all Keynote Speakers, Reviewers, authors, and Committees, for their effort, guidance, contribution and valuable support. Last but not least, thanks to all lecturers and staffs of the Faculty of Science & Technology, Syarif Hidayatullah Jakarta State Islamic University and Universitas Potensi Utama-Medan and other parties that directly and indirectly make this event successful.

Wa billahi taufiq wal hidaayah.
Wallahul muwaffiq ila aqwamit-tharieq.
Wasalaamu ‘alaykumu warahmatullahi wabarakaatuh.

Husni Teja Sukmana
(Organizing Chair)

COMMITTEES

HONORARY CHAIRS
Lukman Hakim Saefuddin, Religious Affairs Minister of the Republic of Indonesia
Rudi Antara, ICT Minister of the Republic of Indonesia
Dede Rosyada, Rector of State Islamic University of Syarif Hidayatullah Jakarta
Bob Subhan Riza, Potensi Utama University

STEERING COMMITTEE
Rika Rosnelly, Potensi Utama University, Indonesia
Agus Salim, State Islamic University of Syarif Hidayatullah Jakarta, Indonesia
Ismail Khalil, Johannes Kepler University, Austria
Tri Haryanto, IT Best Practice, Indonesia
Djoko Soetarno, Coris, Indonesia
Agus Setiawan, Multimatics, Indonesia
Abdul Wahab Abdul Rahman, International Islamic University Malaysia, Malaysia
Sri Hartati, IndoCEISS, Indonesia Computer Electronic and Instrumentation Support Society
Suryadiputra Liawatimena, IEEE Indonesian Section Computer Society Chapter, Bina Nusantara University
Nur Inayah, State Islamic University of Syarif Hidayatullah Jakarta, Indonesia
Syopiansyah Jaya Putra, State Islamic University of Syarif Hidayatullah Jakarta, Indonesia
Rosiyati Mh Thamrin, STMIK Sepuluh Nopember Jayapura, Indonesia

ORGANIZING COMMITTEE
General Chair
Husni Teja Sukmana (State Islamic University of Syarif Hidayatullah Jakarta, Indonesia)

Local Organizing Committee Chair
Edy Victor Haryanto (Potensi Utama University)

Publication
Yusuf Durachman (State Islamic University of Syarif Hidayatullah Jakarta, Indonesia)

Publicity
Yuditha Ichsani (State Islamic University of Syarif Hidayatullah Jakarta, Indonesia)
Asep Taufik Muharram (State Islamic University of Syarif Hidayatullah Jakarta, Indonesia)
Andi Sanjaya (Potensi Utama University)
Hardianto (Potensi Utama University)

Committee Members
Feri Fahrianto (State Islamic University of Syarif Hidayatullah Jakarta, Indonesia)
Nurul Faizah Rozy (State Islamic University of Syarif Hidayatullah Jakarta, Indonesia)
Siti Ummi Masruroh (State Islamic University of Syarif Hidayatullah Jakarta, Indonesia)
Dewi Khairani (State Islamic University of Syarif Hidayatullah Jakarta, Indonesia)

Fitri Mintarsih (State Islamic University of Syarif Hidayatullah Jakarta, Indonesia)
Arini (State Islamic University of Syarif Hidayatullah Jakarta, Indonesia)
Nia Kumaladewi (State Islamic University of Syarif Hidayatullah Jakarta, Indonesia)
Rofiqoh Dewi (Potensi Utama University)
Charles Bronson (Potensi Utama University)
Linda Wahyuni (Potensi Utama University)
Asbon Hendra (Potensi Utama University)
Iwan Fitrianto (Potensi Utama University)
Muhammad Rusdi Tanjung (Potensi Utama University)
Lili Tanti (Potensi Utama University)
Evri Ekadiansyah (Potensi Utama University)
Soeheri (Potensi Utama University)
Haris (Potensi Utama University)

TECHNICAL PROGRAM COMMITTEE

Chair
Ismail Khalil (Johannes Kepler University, Austria)
M Qomarul Huda (State Islamic University of Syarif Hidayatullah Jakarta, Indonesia)

Track Chair:
A'ang Subiyakto (State Islamic University of Syarif Hidayatullah Jakarta, Indonesia)

Technical Program Committee
Kuncoro Wastuwibowo, IEEE Indonesian Section
Shingo Mabu, Yamaguchi University
Normaziah Abdul Aziz, International Islamic University, Malaysia
Akram M. Zeki, International Islamic University, Malaysia
Abdullah Alkalbani, University of Buraimi, Sultanate of Oman
Suvdaa Batsuuri Affiliasi, School of Engineering and Applied Sciences, National University of Mongolia
Suryadiputra Liawatimena, IEEE Indonesian Section Computer Society Chapter, Bina Nusantara University
Dwiza Riana, STMIK Nusa Mandiri, Indonesia
M Qomarul Huda, State Islamic University of Syarif Hidayatullah Jakarta, Indonesia
Muhammad Izman Herdiansyah, Bina Darma University, Palembang, Indonesia
Iwan Setyawan, Kristen Satya Wacana University, Indonesia
Abdul Syukur, Dian Nuswantoro University, Indonesia
Heru Agus Santoso, Dian Nuswantoro University, Indonesia
Affandy, Dian Nuswantoro University, Indonesia
Andrew Tanny Liem, Klabat University, Indonesia
Muhammad Rusli, The School of Information Management and Computer Engineering (STIKOM) Bali, Indonesia
I Gde Putu Wirarama Wedashwara Wirawan, The School of Information Management and Computer Engineering (STIKOM) Bali, Indonesia
Yudi Agusta, The School of Information Management and Computer Engineering (STIKOM) Bali, Indonesia

Taqwa Hariguna,The College of Information Management and Computer Science (AMIKOM) Purwokerto, Indonesia
M. Suyanto, The College of Information Management and Computer Science (AMIKOM) Yogyakarta, Indonesia
Ema Utami, The College of Information Management and Computer Science (AMIKOM) Yogyakarta, Indonesia
Adam Suhaimi, International Islamic University Malaysia
Nurhayati, Syarif Hidayatullah State Islamic University Jakarta, Indonesia
Amin Anjomshoaa, Vienna University of Technology, Austria
Andrzej Gospodarowicz, Wroclaw University of Economics, Poland
Ankhaa Bayar, National University of Mongolia
Ashraf Elnagar, University of Sharjah
Alimuddin, Tirtayasa University, Indonesia
Abdelkader Hameurlain, Institut de Recherche en Informatique de Toulouse Paul Sabatier University
Aries Susanto HT, State Islamic University of Syarif Hidayatullah Jakarta, Indonesia
Bernardo Nugroho Yahya, Ulsan National Institute of Science and Technology, Korea
Christian Huemer, Vienna University of Technology, Austria
Dieter Kranzlmüller, Ludwig-Maximilians-Universität München, Germany
Dimitris Karagiannis, University of Vienna, Austria
Duong Anh Duc, University of Information Technology, Vietnam
Elly Matul Imah, Sepuluh Nopember Institute of Technology, Indonesia
Fauzan Nurdin, International Islamic University Malaysia
Fadi Aloul, American University of Sharjah
Adel Ali Al-Jumaily, University of Technology Sydney
Lintang Yuniar Banowosari, Gunadarma University, Indonesia
Asad I Khan, Monash University
Bessam Abdulrazak, Université de Sherbrooke
Ferry Preska Wathan, Universitas Kader Bangsa, Indonesia
Ford Lumban Gaol, Bina Nusantara University, Indonesia
Geert Poels, Ghent University, Belgium
Imam Shofi, Syarif Hidayatullah State Islamic University Jakarta, Indonesia
Gerald Quirchmayr, University of Vienna, Austria
Gerhard Budin, University of Vienna, Austria
Gerti Kappel, Vienna University of Technology, Austria
Günter Müller, University of Freiburg, Germany
Günther Pernul, University of Regensburg, Germany
Hadipurnawan Satria, Sriwijaya University, Indonesia
Hamideh Afsarmanesh, University of Amsterdam, Netherlands
Hoang Xuan Dau, Posts and Telecommunications Institute of Technology, Vietnam
Husnayati Hussin, International Islamic University Malaysia
Hyerim Bae, Pusan National University, Korea
Harisno, Bina Nusantara University, Indonesia
Irman Hermadi, Bogor Agricultural Institute, Indonesia
Indra Budi, University of Indonesia, Indonesia
Kudang Boro Seminar, Bogor Agricultural Institute, Indonesia
Marimin, Bogor Agricultural Institute, Indonesia
Jarot Sembodo, Bina Nusantara University, Indonesia
Jaafar Gaber, Universite de Technologie de Belfort-Montbeliard
Josaphat Tetuko Sri Sumantyo, Chiba University, Japan

Jamshid B Mohasefi, Urmia University
Pavel Lozhnikov, Omsk State Technical University Rusia
Josef Küng, Johannes Kepler Universität Linz, Austria
Katsumi Tanaka, Kyoto University, Japan
Key Sun Choi, Korea Advanced Institute of Science and Technology, Korea
Khalid Latif, National University of Sciences and Technolgy, Pakistan
Kamel Karoui, RIADI Laboratory
Lenka Lhotska, Czech Technical University, Czech Republic
Luis M. Camarinha Matos, Universidade Nova de Lisboa, Portugal
Masatoshi Arikawa, University of Tokyo, Japan
Mansoor Ahmed, COMSATS Institute of Information Technology
Media Anugerah Ayu, Universitas Siswa Bangsa Internasional, Indonesia
Mizuho Iwaihara, Faculty of Science and Engineering Waseda University, Japan
Mohamed Alkanhal, King Abdulaziz City for Science and Technology
Mohd Farhan Md Fudzee, Universiti Tun Hussein Onn
Imam Machdi, Institute of Statistics
Narayanan Kulathuramaiyer, Universiti Malaysia Sarawak, Malaysia
Nashrul Hakim, State Islamic University of Syarif Hidayatullah Jakarta, Indonesia
Nguyen Thah Binh, IIASA, Austria
Nguyen Tuan, Vietnam National University, Vietnam
Oky Dwi Nurhayati, Diponegoro University, Indonesia
Prihandoko, Gunadarma University, Indonesia
R. Rizal Isnanto, Diponegoro University, Indonesia
Rizal Broer Bahawares, IEEE Computer Society Member
Robert P. Biuk-Aghai, University of Macau, China
Ahmad Nurul Fajar, Bina Nusantara University, Indonesia
Roslina, International Islamic University Malaysia, Malaysia
Sfenrianto, Bina Nusantara University, Indonesia
Shuaib Karim, Quaid-i-Azam University, Pakistan
Somchai Chatvichienchai, University of Nagasaki, Japan
Sourav S. Bhowmick, Nanyang Technological University, Singapore
Sukrisno. Mardiyanto, Institiute Technology of Bandung, Indonesia
Tarek Sheltami, King Fahd University of Petroleum and Minerals,
Tetsuya Furukawa, University of Kyushu, Japan
Thoai Nam, HCMC University of Technology, Vietnam
Taufik, Bina Nusantara University, Indonesia
Vladimir Marik, Czech Technical University, Czech Republic
Werner Winiwarter, University of Vienna, Austria
Wichian Chutimaskul, King Mongkut's University of Technology Thonburi, Thailand
Wikan Danar Sunindyo, Bandung Institute of Technology, Indonesia
Zainal A Hasibuan, University of Indonesia, Indonesia
Zaheer Khan, University of the West of England
Okfalisa, Syarif Kasim State Islamic University Riau, Indonesia
Elly Matul Imah, The State University of Surabaya, Indonesia
Lee Jeong Bae, Busan University of Foreign Studies, South Korea
Bae Jihye, SunMoon University, South Korea
KyungOh Lee, SunMoon University, South Korea
Kwon Jin Bae, SunMoon University, South Korea
Febiansyah Hidayat, Surya University, Indonesia
Adila A Krisnadhi, Wright State University, United State of Amerika

Muhammad Agni Catur, Sampoerna University, Indonesia
Aries Kusdaryono, Ministry of Communication and Informatics, Indonesia
Samsuryadi Sahmin, Sriwijaya University, Indonesia
M. Fachrurrozi, Sriwijaya University, Indonesia
Teddy Mantoro, Sampoerna University, Indonesia
Qonita Shahab, UX Specialist, Netherlands
Murni Mahmud, International Islamic University, Malaysia
Noor Azurati, University Technologi Malaysia, Malaysia
Azizul Azizan, University Technologi Malaysia, Malaysia
Adamu Ibrahim, University Technologi Malaysia, Malaysia
Kamilia Bin Kamardin, University Technologi Malaysia, Malaysia
Akeem Olowo, University Technologi Malaysia, Malaysia
Sya Azmeela, University Technologi Malaysia, Malaysia
Kusrini, Amikom University, Indonesia
Ema Utami, Amikom University, Indonesia
Kim Jin Mook, Sunmoon University, South Korea
Houari Sabirin, KDDI Research, Inc
Khamis Faraj Alarabi Aljribi, Baniwalid University, Libya
Leon Andretti Abdillah, Bina Darma University, Indonesia
Darmawan Napitupulu, Indonesian Research Institute, Indonesia
Golooba Moses, Islamic University In Uganda
Wendi Usino, Budi Luhur University, Indonesia
Mochamad Wahyudi, Bina Sarana Informatika, Indonesia
Roy Rudolf Huizen, Universitas Sumatera Utara Medan, Indonesia
Wisnu Ananta Kusuma, Bogor Agricultural University
Opim Salim Sitompul, Universitas Sumatera Utara Medan, Indonesia
Purwanto, Udinus Semarang, Indonesia
Yana Aditia Gerhana, UIN Sunan Gunung Djati Bandung, Indonesia
Ali Ramdhani, UIN Sunan Gunung Djati, Indonesia
Agus Rifai, International Islamic University Malaysia
Diyah Puspitaningrum, Bengkulu University, Indonesia
Umar Aditiawarman, International Islamic University Malaysia
Purwanto, Udinus Semarang, Indonesia
Dini Octarina Dwi Handayani, Taylors University
Arief Setyanto, AMIKOM Yogya, Indonesia
Elis Ratna Wulan, UIN Sunan Gunung Djati, Indonesia
Muljono, UDINUS Semarang, Indonesia
Untung Rahardja, STMIK Rahardja, Indonesia
Sri Hartati, Gadjah Mada University, Indonesia
Muharman Lubis, Telkom University, Indonesia
Retantyo Wardoyo, Gadjah Mada University, Indonesia
Insap Santosa, Gadjah Mada University, Indonesia
Arief Setyanto, AMIKOM Yogya, Indonesia
Sunny Arief Sudiro, STMIK Jakarta STI&K, Indonesia
Arief Setyanto, AMIKOM Yogya, Indonesia
Evi Triandini, STIKOM Bali, Indonesia
Rahmat Sembiring, Poltek Medan, Indonesia
Achmad Nizar Hidayanto, UDINUS Semarang, Indonesia
Soetam Rizky Wicaksono, Machung Univesity
Nur Sultan Salahuddin, Gunadarma University, Indonesia

Yaqoob Koondhar, Sindh Agriculture University Tandojam, Pakistan
Ankhbayar Yukhuu, The national university of Mongolia
M. Ary Heryanto, UDINUS Semarang, Indonesia
Arief Fatchul Huda, UIN Sunan Gunung Djati, Indonesia
Very Ronny Palilingan, Universitas Negeri Manado, Indonesia
Mohammad Syafrullah, Budi Luhur University, Indonesia
Meyliana, Bina Nusantara University, Indonesia
Masayu Leylia Khodra, ITB, Indonesia
Heru Susanto, Indonesian Institute of Science, Indonesia
Rifki Sadikin, Indonesian Institute of Science, Indonesia
Muhammad Khusairi Osman, Universiti Teknologi Mara (UiTM) Malaysia
Lili Wulandhari, Bina Nusantara University, Indonesia
Meyliana, Bina Nusantara University, Indonesia
Untung Rahardja, STMIK Rahardja, Indonesia

TABLE OF CONTENT

Implementation of The Naïve Bayes Algorithm with Feature Selection using Genetic Algorithm for Sentiment Review Analysis of Fashion Online Companies

Siti Ernawati
STMIK Nusa Mandiri Jakarta
rna2103@gmail.com

Eka Rini Yulia
STMIK Nusa Mandiri Jakarta
ekariniyulia@gmail.com

Frieyadie
STMIK Nusa Mandiri Jakarta
frieyadie@nusamandiri.ac.id

Samudi
STMIK Nusa Mandiri Jakarta
samudi.net@gmail.com

Abstract- Opinion rivalry that occurs in social media have an important role in increasing the potential customers to the company or agency. The review is a rich and useful resource for marketing, social and others for excavations and mining opinions such as views, moods, and behavior. The reviews describe perceptions of something, such as review of a product, review of airline services, reviews of restaurant and others. The analysis of sentiment is an ongoing field of text-based research. The analysis of sentiment or opinion mining is the study of ways to solve problems of public opinion, attitudes, and emotions of an entity, in which the entity may represent individuals, events or topics. Sentiment analysis is an important tool for analyzing opinions in social media. This measurement begins with pre-processing consisting of tokenizing, stopwords removal and stemming. This study uses naïve Bayes algorithm and genetic algorithms as applied feature selection. Selection features aim to classify text for the review of online fashion companies. This measurement results in the classification of text in form of positive text and negative text. Measurements are based on the accuracy of naïve Bayes before addition of genetic algorithms and after addition of genetic algorithms as feature selection. Validation using 10 fold cross-validation. For measurement accuracy using confusion matrix and ROC curve. The purpose of the study is to calculate the increased accuracy of naïve Bayes algorithm if using genetic algorithms for feature selection. The results showed that the genetic algorithm was able to improve the accuracy.

KeyWord—sentiment analisys, fashion online companies, text classification, genetic algorithm, naïve bayes

I. INTRODUCTION

The review is a rich and useful resource for marketing, social and others for excavations and mining opinions such as views, moods, and behavior. The reviews describe perceptions of something, such as review of the product, review of airline services, reviews of restaurant and others. A review can describe the views, attitudes or nature of someone about something. Review available on the internet we can use to be processed in order to produce a knowledge and useful information. The available reviews are a very useful resource in various fields, such as marketing, social and others[1].

Some of the studies that have been conducted on the review include, Analysis sentiment for restaurant reviews using naïve Bayes algorithm [2]. The main problem in text classification is the high dimension of the feature space, this is often the case with text that has tens of thousands of features. Most of these features are irrelevant and not used for text classification can even reduce accuracy and a high number of features can slow down the classification process or even make some classifiers inapplicable [3].

Many consumers are expressing their experiences through social media such as facebook, twitter or other media sites. An online sales company review is a channel that connects consumers to another, they can express opinions about the company in which they have made a purchase transaction.

The purpose of the study is to calculate the increased accuracy of naïve Bayes if using genetic algorithms for feature selection.

II. LITERATURE

A. Sentiment Analysis

Detection of text sentiment has attracted much attention and has grown rapidly in recent years, due to the increased availability of online reviews in digital form. The review is a rich and useful resource for marketing, social and others for excavations and mining opinions such as views, moods, and behavior. For example, whether a review of a positive or negative product, how the mood among bloggers at that time, how the public reflects on political affairs [1].

The analysis of sentiment is an ongoing field of text-based research. The analysis of sentiment or opinion mining is the study of ways to solve problems of public opinion, attitudes, and emotions of an entity, in which the entity may represent individuals, events or topics [4].

Consumer reviews affect whether or not an online sales company is good. Internet becomes an important part of the life. Now, not only from family and friends but also from foreigners located all over the world who may have used

products, shop online on sites, visit places or destinations and see movies can pour their opinions online.

B. Pre-Processing

If the data has been structured data and a numeric value, the data can be presented as a source of data that can be processed further.

The processes performed in pre-processing are:

1) Tokenization
   Tokenize used to separate words or letters of punctuation marks and symbols.
2) Stopwords Removal
   Remove words that are considered unnecessary in the processing of data, for example if, the, of, or, etc.
3) Steaming
   The process of converting a word into a basic word. This method of converting word forms into basic words adjusts the language structure used in the steaming process.

C. Naïve Bayes

Naïve Bayes is an algorithm that is often used in text categorization. The basic idea is to combine the probability of words and categories to estimate the probability of the category of a document [5]. Naïve Bayes is an approach that leads to Bayes theorem, combining previous knowledge with new knowledge. So this is one of the simplified classification algorithms but has high accuracy. [6]

Bayesian Classification is based on the Bayes theorem that has similar classification capabilities to the decision tree and neural network. Bayesian Classification is proven to have high accuracy and speed when applied to databases with large data.

D. Genetic Algorithm

The Genetic Algorithm is one of the optimization algorithms, which was created to mimic some of the processes observed in natural evolution. The Genetic Algorithm is also a strong stochastic algorithm based on the principles of natural and natural genetic selection that is quite successfully applied in machine learning and optimization problems. [7]

The success of the Genetic Algorithm is highly dependent on two factors, population diversity, and selective pressure. There is a strong influence between these two factors. An increase in selection pressure can increase the number of chromosomes directly copied from the previous generation. In contrast, an increase in population diversity can decrease the proportion of inherited chromosomes and lose the opportunity for them to evolve according to offspring. [8]

E. Validation and Evaluation

Confusion matrix provides the decisions obtained in the transfers and testing, the Confusion matrix provides an assessment of the classification performance by object correctly or falsely [9]. Confusion matrix contains the actual information and predictions on the classification system.

Equation model of Confusion Matrix:

1) Accuracy value is the proportion of the number of correct predictions.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (1)$$

2) Sensitivity was used to compare the proportion of TP to positive tuples.

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

3) Specificity was used to compare the proportion of TN to negative tuples.

$$Specificity = \frac{TP}{TN+FP} \quad (3)$$

4) PPV (positive predictive value) is the proportion of cases with a positive diagnosis.

$$PPV = \frac{TP}{TP+FP} \quad (4)$$

5) NPV (negative predictive value) is the proportion of cases with a negative diagnosis.

$$NPV = \frac{TN}{TN+FN} \quad (5)$$

The ROC curve shows accuracy and compares the classification visually. The ROC curve expresses the confusion matrix. ROC is a two-dimensional graph with false positive as horizontal and true positive lines as vertical lines.

Guidelines for classifying the accuracy using the AUC: [9].

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. METHOD

The following are the steps of the research method:

1. Data Collection
   This study uses data taken from websites that provide online reviews. Many of the reviews available from the site include customer reviews of online fashion companies. The data used in this study as many as 200 data consisting of 100 positive reviews and 100 negative reviews.

2. Initial Data Processing
   The next stage is the initial data processing. Dataset used as many as 200 data, 100 positive reviews and 100 negative reviews are used as data training. This dataset in the preprocessing stage must go through three processes.

The processes are tokenization, stop word removal, and stemming.

3. Proposed Model
   The model that researchers propose is to use the feature selection method of the genetic algorithm. Genetic algorithms are used so that the accuracy of using naïve Bayes may increase. The picture below illustrates the model proposed in this study.

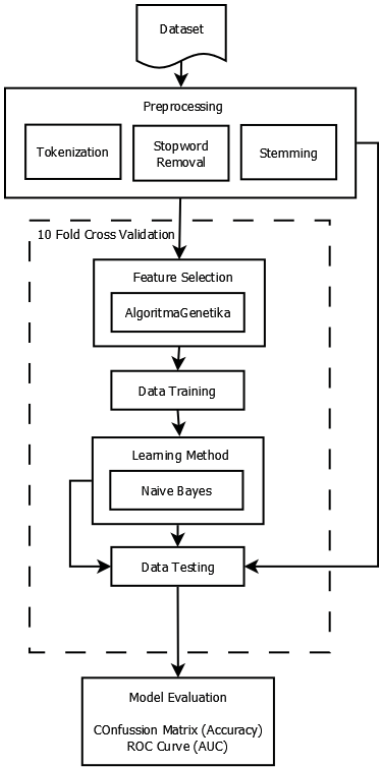

Fig. 1. The Model Proposed

4. Experiments and Testing Models
   a. Setting up datasets for experiments
   b. Input reviews that have not been previously classified
   c. if the text has been inputted all then do pre-process
   d. Design the Naïve Bayes algorithm architecture and do the training and testing and record the accuracy and AUC.
   e. Perform testing with 10 fold cross-validation and look for the value of feature selection.
   f. Designing the naïve Bayes algorithm architecture, the feature selection algorithm is the genetic algorithm and performs.
   g. Training and testing and record the accuracy and AUC.
   h. Perform parameter optimization on genetic algorithm to find out the highest accuracy and AUC.

5. Evaluation and Validation of Results
   The final stage will evaluate the previously tested data by evaluating the comparison results of the whole experiment between using naïve Bayes algorithm with naïve Bayes algorithm and genetic algorithm. The higher value of accuracy, indicating the proposed model is the best.

IV. RESEARCH RESULTS

A. *Model by Classification Method Using Naïve Bayes*

In the study show 10 data from a total of 200 data. 5 words related to the sentiment and most often appears that recommend, disappoint, horrible, good and great.

Validation used 10-fold cross validation for model testing, where each section will be randomly generated. Principle 10-fold cross validation is 1: 9, 1 part becomes data testing and other data into training data, so that 10 part is the chance to be data testing.

TABLE I
Accuracy results using Naïve Bayes algorithm

Accuracy: 68.50% +/- 4.50% (mikro: 68.50%)			
	true negatif	**true positif**	**class precision**
pred.negatif	78	41	65.55%
pred.positif	22	59	72.84%
class recall	78.00%	59.00%	

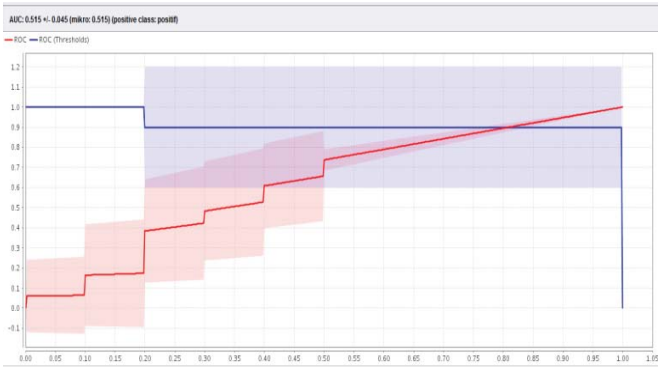

Fig. 2. Graph Area Under Curve (AUC) using Naïve Bayes Algorithm

B. *Model with Classification Method Using Naïve Bayes and Selection of Genetic Algorithm Features*

The optimal parameters in Genetic Algorithm were obtained with population size 50, the number of generation 30, p crossover 0.8 and p mutation 0.08 [10].

To get the highest accuracy results required parameters that require adjustment. Here are the parameters that are adjusted.

TABLE 2
Experiment Plan

Maximum Number of Generation	Population Size	P Crossover	P Mutation	Accuracy	AUC
30-100	5-50	0.5-1.0	0.5-1.0	?	?

TABLE 3
Experimental Results

Maximum Number of Generation	*Population size*	*P Crossover*	*P Mutation*	*Accuracy*	AUC
40	**45**	**0.5**	**0.5**	**87.50%**	**0.819**
40	45	0.5	0.6	87.50%	0.819
40	45	0.5	0.7	87.50%	0.819
40	45	0.5	0.8	87.50%	0.819
40	45	0.5	0.9	87.50%	0.819
40	45	0.5	1.0	87.50%	0.819

The final adjustment of p mutation parameter starting from 0.5-1.0 and there was no change in the value of accuracy and AUC. Therefore, the adjustment of p mutation is taken from the default value of 0.5. From the experimental process that has been done, it can be concluded that to obtain the highest accuracy and AUC value, the optimal parameters for the maximum number of generation values are 40, population size 45, p crossover 0.5 and p mutation 0.5.

TABLE 4
Confusion Matrix Naïve Bayes Algorithm after added selection features of Genetic Algorithm

Accuracy: 87.50% +/- 7.50% (mikro: 87.50%)			
	true negatif	true positif	class precision
pred.negatif	81	6	93.10%
pred.positif	19	94	83.19%
class recall	81,00%	94,00%	

Accuracy value of confusion matrix naïve bayes algorithm after the addition of genetic algorithm are:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (6)$$

$$Accuracy = \frac{(81+94)}{(81+6+94+19)} = \frac{175}{200} = 0{,}875 = 87{,}50\% \quad (7)$$

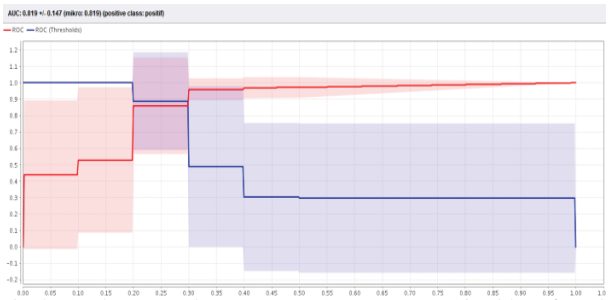

Fig. 3. Graph Area Under Curve (AUC) Naïve Bayes Algorithm after addition selection feature of Genetic Algorithm

TABLE 5
Naïve Bayes Algorithm model before and after using feature selection

	Naïve Bayes Algorithm	**Naïve Bayes Algorithm + Genetic Algorithm**
Successful classification of positive reviews	59	94
Successful prediction of negative reviews	78	81
Accuracy	68.50%	87.50%
AUC	0.515	0.819

Based on results of the above evaluation is known that naïve Bayes algorithm after addition of genetic algorithm feature selection can increase the accuracy value for online fashion company review. Figure 4 is a graph showing the accuracy of the naïve Bayes algorithm and the naïve Bayes algorithm and the genetic algorithm. Figure 5 is a graph showing a value of AUC.

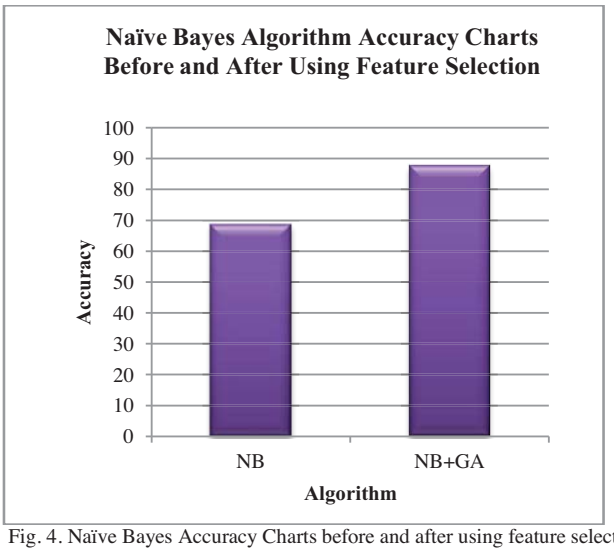

Fig. 4. Naïve Bayes Accuracy Charts before and after using feature selection

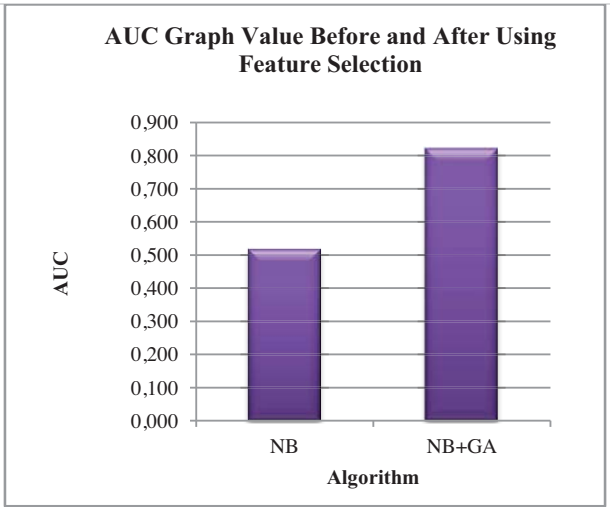

Fig. 5. AUC Graph Value before and after using feature selection

CONCLUSION

Based on the data processing that has been done, merging the naïve Bayes algorithm with selection features of genetic algorithm can improve the accuracy. Online fashion company review can be classified well into positive and negative reviews. Accuracy naïve Bayes algorithm before using feature selection of 68.50% and AUC 0.515. While accuracy after using genetic algorithm feature selection of 87.50% and AUC 0.819. Accuracy increased in the amount of 19.00% and the accuracy of testing using AUC included in Good Classification category.

REFERENCES

[1] H. Tang, S. Tan, and X. Cheng, "Expert Systems with Applications A survey on sentiment detection of reviews," *Expert Syst. Appl.*, vol. 36, no. 7, pp. 10760–10773, 2009.

[2] H. Kang, S. J. Yoo, and D. Han, "Senti-lexicon and improved Na??ve Bayes algorithms for sentiment analysis of restaurant reviews," *Expert Syst. Appl.*, vol. 39, no. 5, pp. 6000–6010, 2012.

[3] J. Chen, H. Huang, S. Tian, and Y. Qu, "Feature selection for text classification with Na??ve Bayes," *Expert Syst. Appl.*, vol. 36, no. 3 PART 1, pp. 5432–5435, 2009.

[4] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.

[5] Z. Zhang, Q. Ye, Z. Zhang, and Y. Li, "Sentiment classification of Internet restaurant reviews written in Cantonese," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7674–7682, 2011.

[6] S. F. Rodiyansyah and E. Winarko, "Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification," *FMIPA UGM*, vol. 6, no. 1, pp. 91–100, 2012.

[7] P. Guo, X. Wang, and Y. Han, "The Enhanced Genetic Algorithms for the Optimization Design," no. Bmei, pp. 2990–2994, 2010.

[8] W. Song, C. H. Li, and S. C. Park, "Expert Systems with Applications Genetic algorithm for text clustering using ontology and evaluating the validity of various semantic similarity measures," *Expert Syst. Appl.*, vol. 36, no. 5, pp. 9095–9104, 2009.

[9] F. Gorunescu, *Data mining: concepts and techniques*. 2011.

[10] S. Günal, "Hybrid feature selection for text classification," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 20, no. SUPPL.2, pp. 1296–1311, 2012.