VARIOUS EXPERIMENTS OF NAVE BAYES AND DECISION TREE ON E-COMMERCE APPLICATION REVIEWS

TUGAS AKHIR

Shelviyanah
41517010008

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2021

VARIOUS EXPERIMENTS OF NAVE BAYES AND DECISION TREE ON E-COMMERCE APPLICATION REVIEWS

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:
Shelviyanah
41517010008

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2021

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:
NIM : 41617010008
Nama : Shelviyanah
Judul Tugas Akhir : Various Experiments Of Nave Bayes And Decision Tree On E-Commerce Application Reviews

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 12 Juni 2021

Shelviyanah

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa	:	Shelviyanah
NIM	:	41517010008
Judul Tugas Akhir	:	Various Experiments Of Nave Bayes And Decision Tree On E-Commerce Application Reviews

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 12 Juni 2021

Shelviyanah

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa	:	Shelviyanah
NIM	:	41517010008
Judul Tugas Akhir	:	Various Experiments Of Nave Bayes And Decision Tree On E-Commerce Application Reviews

Menyatakan bahwa:

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis		Status	
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi		Diajukan	√
		Jurnal Nasional Terakreditasi	√		
		Jurnal International Tidak Bereputasi		Diterima	
		Jurnal International Bereputasi			
	Disubmit/dipublikasikan di :	Nama Jurnal	: SINERGI		
		ISSN	: eISSN : 24601217 | pISSN : 14102331		
		Link Jurnal	: https://publikasi.mercubuana.ac.id/index.php/sinergi/index		
		Link File Jurnal Jika Sudah di Publish	:		

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 12 Juni 2021

Shelviyanah

LEMBAR PERSETUJUAN PENGUJI

NIM	:	41517010008
Nama	:	Shelviyanah
Judul Tugas Akhir	:	Various Experiments Of Nave Bayes And Decision Tree On E-Commerce Application Reviews

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 Juli 2021

(Ir. Emil R. Kaburuan, Ph. D., IPM)

LEMBAR PERSETUJUAN PENGUJI

NIM	:	41517010008
Nama	:	Shelviyanah
Judul Tugas Akhir	:	Various Experiments Of Nave Bayes And Decision Tree On E-Commerce Application Reviews

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 Juli 2021

(Dr. Nenden Siti Fatonah, M.Kom)

LEMBAR PERSETUJUAN PENGUJI

NIM	:	41517010008
Nama	:	Shelviyanah
Judul Tugas Akhir	:	Various Experiments Of Nave Bayes And Decision Tree On E-Commerce Application Reviews

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 Juli 2021

(Vina Ayumi, S.Kom., M.Kom)

LEMBAR PENGESAHAN

NIM	:	41517010008
Nama	:	Shelviyanah
Judul Tugas Akhir	:	Various Experiments Of Nave Bayes And Decision Tree On E-Commerce Application Reviews

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 26 Juli 2021

Menyetujui,

(Dr. Mujiono Sadikin, MT. CISA, CGEIT)
Dosen Pembimbing

Mengetahui,

(Wawan Gunawan, S.Kom, MT)	(Ir. Emil R. Kaburuan, Ph.D., IPM.)
Koord. Tugas Akhir Teknik Informatika	Ka. Prodi Teknik Informatika

KATA PENGANTAR

Assalamu'alaikum Wr.Wb

Puji syukur saya panjatkan kehadirat Allah SWT, yang senantiasa telah melimpahkan rahmat, hidayah dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan Tugas Akhir yang berjudul "Various Experiments Of Nave Bayes And Decision Tree On E-Commerce Application Reviews" ini dengan baik dan tepat pada waktunya. Tujuan dari penulisan laporan Tugas Akhir ini dibuat sebagai salah satu syarat untuk dinyatakan LULUS sebagai sarjana Ilmu Komputer dari Universitas Mercu Buana. Saya menyadari bahwa penyusunan Tugas Akhir ini tidak lepas tanpa bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Kedua orang tua Ayah(Alm) dan Ibu, serta Kakak dan Adik yang tidak pernah lelah untuk senantiasa memberikan motivasi, semangat, doa, dukungan dan kepercayaan kepada saya, sehingga dapat menyelesaikan kuliah dengan baik.
2. Bapak Yahya Sudarya Triana, M. Kom., Ph.D selaku Dekan Fakultas Ilmu Komputer, Program Studi Teknik Informatika Universitas Mercu Buana.
3. Bapak Dr. Mujiono Sadikin, MT, selaku Pembimbing Tugas Akhir yang selalu semangat memberikan waktunya untuk membimbing, memberi arahan dan memberi saran di setiap proses pembuatan Tugas Akhir ini.
4. Bapak Hery Derajad Wijaya, S.Kom, MM selaku Kaprodi Teknik Informatika Universitas Mercu Buana.
5. Bapak Drs. Ahmad Kodar, MT, selaku Pembimbing Akademik yang telah membimbing dan membantu dalam memberikan informasi dan penjelasan mengenai proses akademik selama perkuliahan berlangsung.

6. Bapak Diky Firdaus, S. Kom, MM, selaku koordinasi Tugas Akhir Fakultas Ilmu Komputer, Program Studi Teknik Informatika Universitas Mercu Buana.
7. Seluruh Dosen Program Studi Teknik Informatika yang telah memberikan ilmu, wawasan, dan pengalaman yang bermanfaat selama perkuliahan berlangsung. Memberi kesempatan untuk belajar, berkarya dan juga berkembang.
8. Seluruh Staff Administrasi dan Tata Usaha yang telah banyak membantu dan memberikan kemudahan atas semua pelayanan dan arahannya.
9. Kepada sahabat terbaik saya Eka Ayu Ernawati, Annisa Rizki Liliandari, Wulandari Titi Mawarni, Nazia Fitri Kusubandiah dan Wijanarko Haminoto yang telah membantu meringankan beban dalam menyelesaikan Tugas Akhir ini, serta selalu memberikan dukungan baik itu doa maupun semangat yang tidak pernah henti.
10. Sahabat-sahabat dan rekan-rekan mahasiswa/i program studi S1 Teknik Informatika khususnya teman-teman seperjuangan angkatan 2017 dan perempuan-perempuan Teknik Informatika atas kekompakan, kebersamaan, pertemananya, saling menyemangati, saling membantu, saling mengingatkan, dan saling mendoakan yang sudah terjalin hampir 4 tahun ini. Saya harap pertemanan ini tidak akan terputus untuk selamanya dan kami dapat menyelesaikan Tugas Akhir ini bersama dengan baik.
11. Keluarga besar Teknik Informatika yang telah memberikan semangat dan motivasi selama menjalani perkuliahan.
12. Semua pihak dan personal yang tidak dapat disebutkan satu per satu yang terlibat dalam pembuatan Tugas Akhir ini sehingga dapat selesai dengan baik.

Hasil Tugas Akhir ini masih jauh dari sempurna. Masih terdapat kekurangan dalam penelitian, cara penjelasan dan kekeliruan dalam penulisan. Untuk itu, kritik dan saran dari pembaca sangat dihargai dan diharapkan. Akhir kata, semoga Tugas Akhir ini dapat memberikan manfaat bagi para pembaca.

Jakarta, 12 Juni 2021
Shelviyanah

DAFTAR ISI

SYNERGY Vol. xx, No. x, February 20xx: xxx-xxx
http://publikasi.mercubuana.ac.id/index.php/sinergi http://doi.org/10.22441/sinergi.xxxx.x.xxx

NASKAH JURNAL

VARIOUS EXPERIMENTS OF NAVE BAYES AND DECISION TREE ON E-COMMERCE APPLICATION REVIEWS

Shelviyanah*, Mujiono Sadikin[1]
[1]Department of Informatic Engineering, Faculty of Computer Science, Universitas Mercu Buana
[2]Department of Informatic Engineering, Faculty of Computer Science, Universitas Mercu Buana

Abstract
Currently, there are many kinds of applications that offer e-commerce services, examples of e-commerce applications are Tokopedia and JD.ID. Many users who use the application and provide reviews of the application are challenging to analyze. This research is focused on text classification of user reviews of the Tokopedia and JD.ID applications by comparing the performance of the Naïve Bayes and Decision Tree methods and measuring the effect of various experiments to find the method that has the best assessment results and performance with correct predictions in classifying positive or negative reviews. In conclusion, the tuning parameters, the use of N-gram, and the selection of Chi-Square features have an effect on several cases in the performance of the two models. The results showed that the Decision Tree model with a proportion of 80:20 was superior to the Naïve Bayes model with an accuracy of 93.50% while the Naïve Bayes model only had an accuracy of 91%.

Keywords:
Sentiment analysis;
Text classification;
E-commerce;
N-grams;
Feature selection

Article History:
Received: May 2, 2019
Revised: May 29, 2019
Accepted: June 2, 2019
Published: June 2, 2019

Corresponding Author:
Shelviyanah
Informatic Engineering Department, Mercu Buana University, Indonesia
Email: 41517010008@student.mercubuana.ac.id

INTRODUCTION

The development of online business in Indonesia is now very fast, one of which is by doing online shopping. In 2021it is estimated that the value of e-commerce transactions will increase by more than 230 percent to US$ 4.48 trillion or equivalent to Rp. 60,467 trillion from the 2014 position which only reached US$ 1.8 trillion [1]. Online shopping is a way of shopping through electronic communication tools or social networks that are used in buying and selling transactions, is a popular activity today where buyers do not need to come to the store just to buy the desired item, the price of goods purchased is also cheaper than if bought conventionally. Now it is enough to go through a digital e-commerce platform to order the desired goods without knowing regional boundaries, then payment can be made through various transactions such as through bank accounts, through electronic wallets, through minimarkets, and some online stores also provide payments via COD (Cash on Delivery). After making the transaction, the goods are ready to be sent by the online store to the consumer's address.

At this time there are various applications that offer e-commerce services. One example is the JD.ID and Tokopedia applications. The Tokopedia application has the second largest number of visitors after the Shopee application in the fourth quarter of 2020 with a total of 114.7 million visitors. Meanwhile, JD.ID is in 9th place with 4.2 million visitors [1]. Tokopedia and JD.ID are available and can be downloaded through the Google Play Store application. During the use of e-commerce applications, of course, users not only get comfort and convenience in the use or services provided but can also

experience various kinds of obstacles that may occur at any time.

The Google Play Store application provides a review feature to assist users in submitting reviews about problems that occur or dissatisfaction with the services they use. In addition to user reviews can also provide suggestions for future progress of the application. The large number of users who download the Tokopedia and JD.ID applications and the many reviews submitted by users every day, it is not easy to do a manual assessment. Therefore, it can be overcomeby conducting sentiment analysis research. Sentiment analysis [2] is a process of using text analytics to obtain various data sources through several types of social media platforms and the internet and is one of the fields of NLP known as natural language processing.

The problem that occurs when classifying sentiment analysis on text is the number of attributes used in the dataset. Generally, the attributes obtained from text sentiment are very large and not all data has a major influence in making the classification function, and it is possible to reduce accuracy results [3]. To overcome this problem, we need a method by performing feature selection as has been done by several previous research [4, 5, 6, 7] on the classification of text documents. With this system, it will be possible to help the company Tokopedia and JD.ID in evaluating the advantages and disadvantages, knowing user desires, and improving the quality of the application, so that users will feel comfortable using the application in doing online shopping.

There have been several previous studies on opinion analysis and data analysis with related classification methods carried out by researchers. Irma et al. [8] predicting the rating on beauty reviews using MNB and N-grams, it was found that the model with the use of complete preprocessing and the combination of N-grams resulted in better accuracy reaching 97% on the tolerance test 1 and 96% on the sentiment review test. In their research, Desi and Umniy [9] classified text documents as complaints or non-complaints in the marine and fisheries domain using the Random Forest algorithm and applying parameter settings to obtain the best model performance of 95%.

Aulya et al. [10] conducted an analysis to determine public sentiment towards the government's New Normal policy. Obtained the highest accuracy of the model using the parameters of the Naive Bayes algorithm with the N-Gram Trigram type is 84%, there are also Precision values obtained at 84%, Recall at 86%, and f1-Score at 85%. Mujiono and Fahri [11] predict the potential risk on customer credit datasets using the C4.5 and NB algorithms. The results show that the C4.5 model is more suitable because it produces an accuracy of 83.33% than NB which is based on the conditional probability of the input variable and the accuracy achieved is only 80.67%.

This research is focused on text classification of Tokopedia and JD.ID user review data sourced from the Google Play Store application by comparing the performance of the Naïve Bayes and Decision Tree methods and measuring the effect of various experiments used to find the method that has the best accuracy and performance results. With precise predictions in classifying texts as positive or negative reviews review. This study consists of 4 parts, namely the introduction (Part 1), the research method (Part 2), the results and discussion (Part 3), and finally the conclusion (Part 4).

METHOD

This study used quantitative research methods and experimental research. The initial stage of this research begins with preparing the dataset. The dataset used in this study is a review dataset of the Tokopedia online shopping application and a review of the JD.ID online shopping application taken from the Google Play Store website. The dataset obtained is then carried out in the data labeling stage based on the rating of each review. After the data labeling stage is carried out, proceed to the pre-processing stage used in this study, namely case folding, tokenizing, convert *slang words*, negation handling, stemming, and stopword removal.

After doing the pre-processing stage and obtaining clean and structured data that is ready to be processed, the next step is to divide the data into training data and test data. The distribution of the proportion of the data is also carried out with various experiments to see which proportion of the data produces good accuracy for the classification between 90:10, 80:20, and 70:30. After the distribution of the dataset has been carried out, proceed with the feature extraction process. In this study using the TfidfVectorizer feature extraction by applying n-grams carried out with various n-gram experiments to see which results in classification accuracy both between unigrams (1.1), combinations of unigrams and bigrams (1,2), as well as combinations of unigrams and trigrams (1,3).

After the feature extraction process, the training data in this study were used for the learning process on the proposed algorithm, namely Naïve Bayes and Decision Tree, so that the proposed classification model was formed. Naïve Bayes and Decision Tree classifications are carried out using the grid search method as well as validation with cross validation k=10 so as to obtain the optimal parameter values that will be given to test the model. The next step is to test the Naïve Bayes and Decision Tree classification models using test data. Based on the model test, the analysis can be done by calculating the accuracy obtained using*confusion matrix*. In this study also conducted experiments, namely comparing the accuracy results of the proposed algorithm method without a grid search process, N-grams, and feature selection with the proposed algorithm method by applying grid search, N-grams, and feature selection to see the comparison of these experiments. Which is suitable and produces good accuracy for the classification of the dataset in this study.

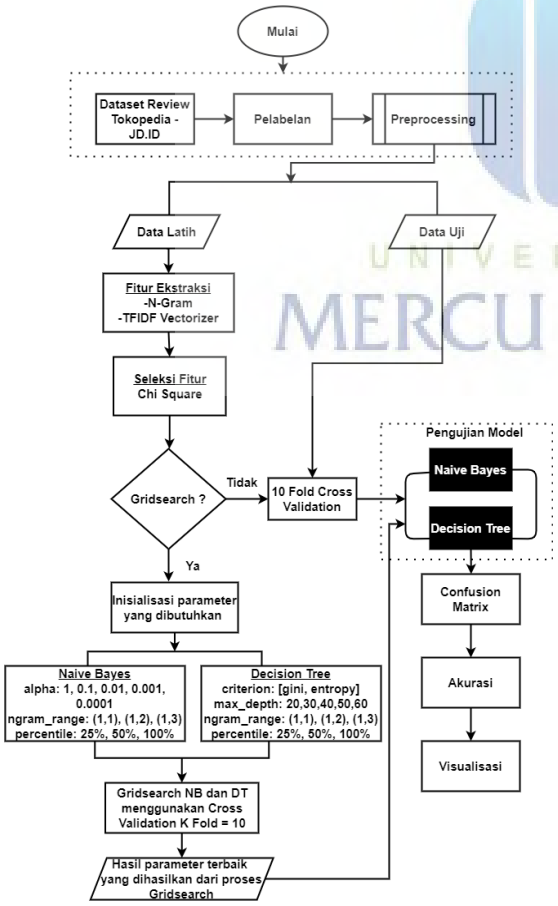

Figure 1. Flowchart of text classification

Data collection

In this study, the dataset used was sourced from user reviews of the Tokopedia and JD.ID online shopping applications in the Google Play Store application. The review data was selected based on the reviews received from January to March 2021. The collection and retrieval of review data was carried out using a package that was already available in the Python programming language, namely google-play-scraper.

The package has provided an API that can be used to perform data retrieval on the Google Play Store application. The review data taken by the author is the data that is considered the most relevant and the latest data using the newest and most.relevant functions and the data is in Indonesian form. The data obtained are 4000 user reviews. In addition to reviewing each user, the author also retrieves information such as username, date, as well as the rating of each review of each user. The data that has been taken is data that has not been carried out in stages *pre-processing* can be seen in Table 1.

Table 1. Tokopedia Review Data Results

Review	Rating
Great!! Tokped has a lot of bugs, if you want to go to OVO via tokped first (pdhl my cellphone is capable of heavy applications), want to create a new account, you can't even use fb, there's always a lot of errors :) Dare to invite BTS + Blackpink, but the performance of the crew ambyar, raise the price of the staff! Let the maximum work£	1
Tokopedia is very helpful in shopping, especially during the current pandemic, you don't need to leave the house but you can still shop. I hope that in the future this application can develop to be even better...	5
JD.ID really helped me to get Products anythingthe original. The delivery is also very fast. Good luck to JD.D....	5
Disappointed and tired of buying electronics on an ID, broken ac, forced to sell, unable to return, for various reasons, for those who want to buy electronics as ID, think again, because if you receive damaged goods, you can't return them, even though the price is 3 million to the top, we are the ones who are responsible for our own loss	1

Data Labeling

The labeling of the review sentiment analysis in this study was carried out using a lexicon dictionary. The lexicon dictionary used is sourced from the github repository [21]. This labeling is done based on the sum of the weight values obtained, the weight in labeling is for positive words worth 1, negative words are given a value of -1 while words that are not in the lexicon dictionary are given a value of 0, if the weight value obtained is > 1 then it will be

labeled positive while if the weight value obtained 0 then it will be labeled negative. The results show that positive labels are greater in number than negative labels. Positives data are 2332 negatives data are 1668.

Data Preprocessing

Before carrying out the classification process for sentiment analysis or text mining, the data needs to be prepared in advance to be processed, because the overall results of the review data obtained consist of many unstructured words such as abbreviations, emoticons, symbols, urls or links, numbers, and the use of slang or slang vocabulary that is not in accordance with KBBI standards. This process is known as pre-processing technique. Pre-processing is an early stage that is useful for making data more structured. This process will greatly affect the performance of the classification algorithm. This research consists of 7 stages of pre-processing, which are as follows.

1. *Case Folding*: is the process of changing all letters into the same form, namely lower cases to minimize variations in writing text data.
2. *Cleansing*: is a process that is carried out to eliminate all spec characterssial such as (@,#,link), numeric characters, punctuation marks (such as periods, commas, question marks, exclamation points, etc.), blank characters (excessive spaces or spacing), emoji characters, repetitive characters.
3. *Tokenizing*: is a process of dividing the text in each document into a series of tokens or terms. The results are each the word in the text is represented as a token or term.
4. Conversion *Slang Word*: is a process of changing words in non-standard form into words in standard form in accordance with KBBI standards, because the review data obtained are mostly filled with the use of foreign words or known as slang.
5. Negation: is a process to overcome the problem of negation of a word, negation greatly affects the polarity of other words. Illustration of the process of negation, example sentence "This application is not good, bugs often occur" after negation will be "this application" not good, bugs often occur".
6. *Stemming*: is a process of changing words into their basic form, where at this stage each word is identified to be changed into its basic form. This study uses a library provided by Python, namely the Sastrawi library which applies the Nazief and Adriani's Stemmer algorithm for Indonesian stemming needs.
7. *Stopword Removal:* a process that aims to filter all tokens and then eliminate them if the token is in the stopword list. This study uses an Indonesian stopword list from the NLTK (Natural Language Toolkit) library.

Extraction and N-Gram Features

In the feature extraction process or word weighting, the Term Frequency – Inverse Document Frequency (TF-IDF) method is used to obtain the weight value of each word in the data used. TF-IDF is a well-known method for evaluating the importance of a word in a document. The Term Frequency of a given term (TF) is calculated as the number of times the term appears in the document by the number of words in the document. IDF is used to calculate the importance of a term. TF-IDF presents word frequency scores, especially for interesting words, such as words that often appear in one document but not for the whole document [12]. This process is carried out by calculating the weight of each word in the training data using the sklearn TfidfVectorizer library.

In this study, in the feature extraction process, N-grams are applied by solving words in sentences including unigrams, namely word solving with n = 1 or single term and unigram+bigram and unigram+trigram. The following is an illustration of the application of N-gram in the sentence "good shopping, the seller's estimate is disappointed that Tokopedia can't sell discipline":

Unigram: 'good', 'shopping', 'estimated', 'the seller', 'disappointed', 'tokopedia', 'no', 'can', 'disciplined', 'sell'.

Unigram+Bigram: 'good', 'good shopping', 'shopping', 'estimated shopping', 'estimated', 'estimated seller', 'seller', 'disappointed seller', 'disappointed', 'disappointed Tokopedia', ' Tokopedia', 'tokopedia doesn't', 'no', 'can't', 'can', 'can be disciplined', 'disciplined', 'selling discipline'.

Unigram+Trigram: 'good', 'good estimate shopping', 'shopping', 'seller estimate shopping', 'estimated', 'disappointed seller estimate', 'seller', 'disappointed seller at Tokopedia', 'disappointed', 'disappointed' Tokopedia doesn't', 'Tokopedia', 'Tokopedia can't', 'No', 'Can't be disciplined', 'Can', 'Can sell discipline'.

Selection Features

Feature selection works based on the process of reducing irrelevant feature spaces by

minimizing each of these irrelevant attributes [3]. The feature selection method is very diverse, for example, Chi-Square, Variance Threshold, Mutual Information, Extra Tree Classifier, and others. In this study, Chi-Square feature selection is used for the feature selection process.

Chi-Square is a statistical method of testing discrete data hypotheses that evaluates the correlation between two variables and determines whether the variables are unrelated or interrelated. This method uses chi square in feature selection. Chi-square feature selection is done by sorting the features based on the results of the chi square feature selection from high to low value [13]. The formula given can be seen in Equation 1:

$$X^2(t, c)=\frac{N(AD-CB)^2}{(A+C)(B+D)(A+B)(C+D)} \quad (1)$$

A is the number of reviews in class C containing the term (t). B is the number of reviews that are not in class C but contain term (t). C is the number of reviews in class C but does not contain term (t). D is the number of reviews that are not in class C and do not contain the term (t). N is the total of all documents. t is term while C is class or category.

To be able to select unused features based on the chi-square value of a term against a class, a single chi-square value of the term is required. To be able to find out the single chi-square value of a term, it is obtained by adding up the chi-square value of each term between classes [13]. The function to get a single chi-square value for each term can be seen in Equation 2.

$$X^2(t) = \sum x\ 2\ (t, c) \quad (2)$$

After knowing the chi-square value in each term, the terms are sorted based on the largest to the smallest chi-square values. This shows that the greater the chi-square value, the more dependent a feature is and the more important the feature is to use in the classification process [13].

Naïve Bayes Classification

Naive Bayes is a method used for binary classification and multi class. This method is a method based on the Bayes theorem proposed by a British scientist Thomas Bayes with the assumption that there are independent properties between existing attributes or variables, in other words, each attribute or variable that exists does not depend on each other. The basic concept of Bayes' theory is the conditional probability P($^{C_j|\,X}C_j|\,X$), where the value of X is posterior and the value of C is prior. Prior is knowledge about the characteristics of a parameter or experiencein the past, while the posterior is the opposite, namely the expected characteristics of future events [14]. Calculation of Naïve Bayes using Equation 3.

$$P(C_j|X) = \frac{P(X)\ P(C_j)}{P(X_i)} \quad (3)$$

Equation 3 illustrates the training data which is categorized into several class categories, namely Cj = {$C1$, $C2$, Ca, … , Ck }, and the prior probability for each class category is P(Cj), where j = 1,2,3,…..k. As for the symboldi = ($W1$, … , , … Wm) is a word in the document in the form of Wj , where j = 1,2,3,…m, belongs to the class category Cj. To classify documents Xi, is done by calculating the probability value of all documents (posterior probability).

Decision Tree Classification

Decision Tree is a popular classification method because it is widely used practically. This method seeks to find a classification model that is resilient to noise. This method maps the observations of an item so that a conclusion is obtained about the target value of an item which is described in the form of a decision tree. Decision trees are known as classifications based on the shape of the structure, namely the root which is the topmost initial node of a decision tree, the leaf which contains the final decision, and the intermediate node that determines several tests to be carried out on a single attribute value that has one branch and subtree for each the results of the test [15].

Several types of decision tree algorithms, namely ID3 which uses information gain to determine which attributes will be used for classification. C4.5 which is the successor of ID3 which can handle continuous and missing attributes. CART is a dynamic learning method that can produce regression trees and classification trees depending on the dependent variable [15].

Classification Performance Evaluation

Evaluation of the performance of a classification model is generally carried out using a set of test data that is not used in training the classification model with a certain size. There are various measures used in assessing or evaluating the performance of the classification model in this study, namely,

accuracy, precision, recall, and f1-score. To measure the performance evaluation of the classification model, it is seen based on the results of the Confusion Matrix. The confusion matrix is a measure to evaluate how well the classification model performs in recognizing real samples from different classes.

RESULTS AND DISCUSSION

For the whole process from collecting datasets, training, testing models, and evaluating models using Python 3.6. The review dataset used is 4000 reviews. After all datasets have been labeled and pre-processed, it can be seen in Table 2 where the results are in the form of basic words that will be used during the feature extraction process.

Table 2. Pre-processing Results

Raw Data
The only place for shopping that is safe and truly satisfying all branded goods from abroad and domestically out there that is hard to find and has to go around is at Tokopedia, shopping from home while sitting comfortably without having to go around expensive, no problem as long as it's satisfying, it's really according to dreams , the packaging is always neat, never damaged, let alone tricking the sellers, the sellers are friendly, even I have never hesitated to buy, ask to be canceled first, it can be canceled, it's not complicated, like at the shop next door, Thx, keep it up.
Pre-processing data
shopping, safe, satisfied, goods, branded, country, domestic, search, hard, mobile, shopeeshop, home, sit, cute, traveling, expensive, no_problem, satisfying indeed, suitable, dreampacking, neat, never_ever, broken, scam, selling friendly, doubtful, buy, cancel, cancel, no, complicated, rich, shop, sidethx, tokopedia, jaya

After the feature extraction process is carried out, before entering the classification analysis stage, the review data needs to be divided into two parts, namely training data and test data first. In this analysis, several experiments were carried out in dividing the proportion of training data and test data, namely 90% training data with 10% test data, 80% training data with 20% test data, and 70% training data with 30% test data which can be seen in Table 3.

Table 3. Total Training Data and Test Data

Data	90:10	80:20	70:30
Train	3600	3200	2800
Test	400	800	1200

Before proceeding to the model testing stage, it is necessary to perform *tuning parameters* first during the model training process. In determining the combination of tuning parameters and obtaining the best combination of tuning parameters, it is done using the *gridsearchcv* method. This method is carried out by scanning the data to find parameters that provide optimal results for the proposed model by applying the cross validation technique as a performance metric that can prevent overfitting problems with a 10-fold number of folds.

In this study, the process of feature extraction and feature selection can be carried out simultaneously using a pipeline system. Then for the search workflow the combination of the required parameter values is carried out in parallel. For the feature extraction step, the TF-IDF Vectorizer is applied by applying N-gram parameters with a range of (1,1), (1,2), and (1,3). For the feature selection stage, use the chi-square selection feature based on the value of percentile which has been determined, the value of percentile consists of 25%, 50%, and 100%. As for the model, several parameters are given to consider which will be searched for the optimal parameters. Naïve Bayes model uses the alpha parameter with values of 10-1, 10-2, 10-3, 10-4, and 1. Decision Tree model uses criterion and max depth parameters, where the criterion parameters used are "Gini" and "Entropy" , the max depth parameter is 20, 30, 40, 50, and 60. ResultsThe overall search grid can be seen in Table 4-6.

Table 4. Naïve Bayes Gridsearch Results with Chi Square Selection Feature

Percentile	Alpha Param	TF-IDF Vectorizer Score Accuracy (%)		
		(1,1)	(1,2)	(1,3)
25%	1	83.56	82.68	79.56
	0.1	84.46	86.25	85.09
	0.01	84.50	87.28	86.81
	0.001	84.28	87.15	86.96
	0.0001	84.06	86.78	86.96
50%	1	83.43	81.62	78.34
	0.1	85.12	86.25	85.31
	0.01	85.12	87.40	87.15
	0.001	84.68	88.12	87.62
	0.0001	84.43	88.15	87.96
100%	1	82.96	81.25	78.25
	0.1	85.56	87.37	86.96
	0.01	85.71	88.78	88.43
	0.001	85.18	89.03	88.59
	0.0001	84.93	88.78	88.78

From Table 4 above, the results show that the best parameter of the Naïve Bayes model is found in the alpha parameter with a value of 0.001 with an accuracy gain of 93.81%. Where these results are obtained from feature extraction with N-gram 1,2 (combination unigram with bigram) and feature selection with value percentile of 100%. In terms of the alpha parameter, it can be seen that the smaller the

alpha parameter, the better the accuracy obtained. In terms of N-grams, it can be seen that the Naïve Bayes model is more suitable with the use of N-gram which are a combination of unigram and bigram. Meanwhile, in terms of feature selection in the dataset of this study, it does not have an effect on accuracy.

Table 5. Results of the Gridsearch Decision Tree Criterion "Gini" with the Chi Square Selection Feature

percentile	Alpha Param	TF-IDF Vectorizer Score Accuracy (%)		
		(1,1)	(1,2)	(1,3)
25%	20	90.53	89.84	89.62
	30	91.81	91.46	91.93
	40	92.65	92.43	91.87
	50	92.87	92.65	91.93
	60	93.34	92.87	91.90
50%	20	90.75	90.21	90.09
	30	92.09	92.00	91.59
	40	92.28	92.12	92.18
	50	92.78	93.00	92.46
	60	93.15	92.62	92.28
100%	20	90.18	89.78	89.87
	30	91.68	91.56	91.43
	40	92.43	92.09	92.53
	50	92.31	92.62	92.46
	60	92.93	92.62	93.00

Table 6. Results of the Gridsearch Decision Tree Criterion "Entropy" with the Chi Square Selection Feature

Percentile	Alpha Param	TF-IDF Vectorizer Score Accuracy (%)		
		(1,1)	(1,2)	(1,3)
25%	20	90.40	90.21	90.34
	30	91.71	91.74	91.87
	40	92.56	92.46	92.53
	50	92.68	92.84	93.00
	60	92.81	92.75	92.81
50%	20	90.34	90.40	90.25
	30	91.93	91.78	91.90
	40	92.87	92.46	92.59
	50	92.84	93.25	92.84
	60	93.09	92.75	92.68
100%	20	90.78	90.24	89.93
	30	92.12	91.56	91.56
	40	93.00	92.34	92.06
	50	92.68	92.62	92.59
	60	92.93	93.09	92.65

From Table 5-6 above, for the Decision Tree model, the best parameters are the criterion "Gini", max depth with a value of 60 with an accuracy of 93.34%. Where the results are obtained from feature extraction with N-gram 1.1 (unigram) and feature selection with value of percentile of 25%. In terms of parameters, the "Gini" criterion parameter is better than the "Entropy" criterion parameter in terms of accuracy. The difference between Naïve Bayes and Decision Tree models in this study is the use of N-grams and feature selection. As previously explained, feature selection in Naïve Bayes has no effect on accuracy results, while in the Decision Tree model it has an effect on accuracy results. Another difference with the use of N-gram, in the Naïve Bayes model the model works better when applying the N-gram combination of unigram and bigram, while in the Decision Tree model the use of N-gram does not affect the accuracy results because of the accuracy produced when applying N-gram between unigram-bigram combinations and unigram-trigam combinations tend to obtain low accuracy.

After getting the best model from the results of the training process, these results are used in conducting several experimental experiments to get the best model testing results. The experiment was carried out by comparing the accuracy results of the proposed model without a grid search process, N-gram, and feature selection with the proposed model by applying grid search, N-gram, and feature selection to see the comparison of the experiments that matched and resulted in accuracy. Good for classification. The experimental results with 4 different metrics can be seen in Table 7-8.

Table 7. Experimental Results of the Naïve Bayes Model

Model Classification	Accuracy (%)			
	Acc	P	R	FM
No Alpha, N-gram and Chi Square Feature Selection parameters	83.00	83.30	80.35	81.32
Without Alpha parameter and with N-Gram (1.2)	83.50	87.00	79.30	80.96
With N-grams (1.2) and Alpha 0.001	91.00	91.49	89.47	90.28

Table 8. Decision Tree Model Experiment Results

Model Classification	Accuracy (%)			
	acc	P	R	FM
Parameterless, N-Gram, Chi Square Feature Selection	93.12	93.33	92.10	92.64
Without Criterion and Max Depth parameters, with N-Gram (1.1) and Chi Square Feature Selection	92.13	92.25	91.05	91.57
With N-grams (1.1), Criterion Entropy Parameter, Max	93.50	93.51	92.71	93.08

Depth 60 and Chi Square Feature Selection			

From Table 7-8 above, for the Naïve Bayes model, it can be seen that the best model is obtained in the model that uses N-gram and the addition of alpha parameters, while for the Chi-Square feature selection in the Naïve Bayes model it does not affect the accuracy results obtained. This shows that only the addition of tuning parameters and selection of N-gram features can help improve classification performance. The best accuracy obtained is 91.00% with a precision value of 91.49%, recall is 89.47% and f-measure is 90.28%. The increase obtained from experiments without using alpha parameters, without applying N-gram with those using alpha and N-gram parameters reached 7-8%.

For the Decision Tree model, it can be seen that the best model is obtained from a model that applies Chi-Square feature selection and the addition of alpha parameters. This shows that the addition of tuning parameters and feature selection contribute to improving the classification performance. The best accuracy is 93.50% with a precision value of 93.51%, recall is 92.71% and f-measure is 93.08%. The increase obtained from experiments without using alpha parameters, and feature selection using alpha parameters and feature selection only reached 1%. In contrast to the Naïve Bayes model, the accuracy gain from each experiment is quite good.

In addition to experimental testing based on parameters, N-grams, and feature selection were also tested based on the distribution of the data proportions described earlier. The results obtained based on the distribution of the proportion of data in each model can be seen in Table 9.

Table 9. Results of Data Proportion Experiments

Model Classification	Proportion of Data Training and Testing		
	90:10	80:20	70:30
Naive Bayes	88.00	91.00	89.42
Decision Tree	93.30	93.50	91.70

The results from Table 9 show that the average distribution of data proportions with a percentage of 80% of the training dataset and 20% of the testing dataset produces good accuracy in both models compared to the distribution of other data proportions. This experiment shows that the amount that is not too large in the training dataset and not too small in the testing dataset, the better the accuracy of the model obtained.

After the classification process is complete, the model with alpha parameters, application of N-gram, feature selection, and with the proportion of data 80% training data 20% data testing, so the next step is to evaluate the performance of the model with accuracy parameters and confusion matrix. The results of the confusion matrix for each model can be seen in Table 10.

Table 10. Confusion Matrix Results

Classification Model	Original Class	Prediction Class	
		Negative	Positive
Naive Bayes	Negative	255	53
	Positive	19	473
Decision Tree	Negative	275	33
	Positive	19	473

Based on Table 10 the results of the confusion matrix, for example in the Naïve Bayes model from the negative class side, it is classified as true negative as many as 225 times with predicted errors of 53 as positive classes. From the positive class side, it was classified as true positive only 473 times, 19 were predicted as negative class. In the Decision Tree model, from the negative class side, it is classified as true negative 275 times with predicted errors of only 33 as positive classes. From the positive class side, it was classified as true positive only 473 times, 19 were predicted as negative class. From the results of the confusion matrix in each model, it shows that the Naïve Bayes model is better at predicting each class than the Decision Tree model.

Prediction result the qualifications obtained can be seen in Figure 2-4. The overall chart visualization and word cloud, combined from the Tokopedia and JD.ID applications, show that they have a more dominant positive sentiment with a percentage of 63.2% and a negative only 36.8%.

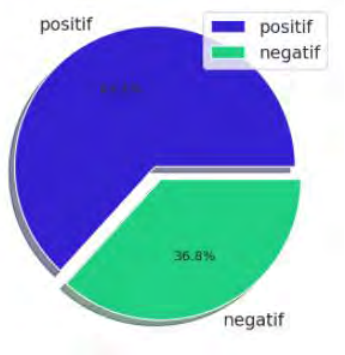

Figure 2. Predicted results of text classification

In Figure 3-4 below, despite the dominant positive sentiment, the Tokopedia and JD.ID applications still have drawbacks based on negative reviews, namely failing to download and verify, incorrectly sent goods, transaction cancellations, lack of system updates, disappointed in the service , and others. As for the advantages that are owned based on positive reviews, namely the application is easy, useful, shopping is fast, shopping is satisfied, delivery is fast, the application is helpful, and others.

Figure 3. Positive prediction wordcloud results

Figure 4. Negative prediction wordcloud results

CONCLUSION

Based on the results of analysis and testing with several experiments on the classification of user review texts on the Tokopedia and JD.ID e-commerce applications, it shows that parameter tuning, the use of N-Gram, and Chi-Square feature selection can have an influence on the performance of the two models. Where there is a difference in this study lies in the use of N-Gram and feature selection. The Naïve Bayes model works better when using N-Gram for feature extraction. As for the Decision Tree model, the use of N-Gram has no effect on the accuracy results. For the Chi-Square feature selection in the Naïve Bayes model, it does not affect the accuracy results obtained. As for the Decision Tree model, it works better with the use of feature selection with a percentage of 25%. This shows the Decision Tree is more suitable with a smaller percentage of the number of features.

The overall results show that the Decision Tree model is superior to the Naïve Bayes model because in Nave Bayes the joint relationship in features reduces algorithm performance, where the acquisition of accuracy, precision, recall, and f-measure values are 93.50%, 93.51%, 92.71%, 93.08% respectively. Then followed by the Naïve Bayes model with 91.00%, 91.49%, 89.47%, 90.28% respectively. For further sentiment analysis research, it is recommended to test with different feature selection, compare more classification algorithms, use other methods such as deep learning or other methods. Thus, the results obtained can help more in obtaining better performance.

ACKNOWLEDGMENT

The author would like to thank the Faculty Leaders Computer Science, Mercu Buana Meruya University, for supporting this research.

REFERENCES

[1] Jarot, D.B. (2021, 11 Februari). 10 E-commerce dengan Pengunjung Terbesar pada Kuartal IV 2020. Available: https://databoks.katadata.co.id/datapublish/2021/02/11/10-e-commerce-dengan-pengunjung-terbesar-pada-kuartal-iv-2020 (diakses tanggal 5 Mei 2021).

[2] E. Sutoyo, and A. Almaarif, “Twitter sentiment analysis of the relocation of Indonesia's capital city,” Bulletin of Electrical Engineering and Informatics, vol.9, pp. 1620-1630, 2020.

[3] F. Es-Sabery et al., "A MapReduce Opinion Mining for COVID-19-Related Tweets Classification Using Enhanced ID3 Decision Tree Classifier," in IEEE Access, vol. 9, pp. 58706-58739, 2021.

[4] H. Syahputra et al. “Setiment Analysis of Public Opinion on The Go-Jek Indonesia Through Twitter Using Algorithm Support Vector Machine.” J. Phys, 2020.

[5] F. Zamachsari, G. V. Saragih, Susafa’ati, and W. Gata, “Analysis of Sentiment of Moving a National Capital with Feature

Selection Naive Bayes Algorithm and Support Vector Machine," Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi), vol *4*, pp. 504 – 512, 2020.

[6] H. M. Keerthi Kumar and B. S. Harish, "A New Feature Selection Method for Sentiment Analysis in Short Text" Journal of Intelligent Systems, vol. 29, 2020.

[7] S. Bahassine, A. Madani, M. Al-Sarem, and M. Kissi, "Feature selection using an improved Chi-square for Arabic text classification," *J. King Saud Univ. Comput. Inf. Sci.,* vol 32, pp. 225-231, 2020.

[8] Nurhayati, A. E. Putra, L. K. Wardhani and Busman, "Chi-Square Feature Selection Effect On Naive Bayes Classifier Algorithm Performance For Sentiment Analysis Document," 2019 7th International Conference on Cyber and IT Service Management (CITSM), pp. 1-7, 2019.

[9] Y. D. Setiyaningrum, A. F. Herdajanti, C. Supriyanto and Muljono, "Classification of Twitter Contents using Chi-Square and K-Nearest Neighbour Algorithm," 2019 International Seminar on Application for Technology of Information and Communication (iSemantic), pp. 1-4, 2019.

[10] Y. Zhai, W. Song, X. Liu, L. Liu and X. Zhao, "A Chi-Square Statistics Based Feature Selection Method in Text Classification," 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS), pp. 160-163, 2018.

[11] B. Venkateswarlu, H. L. Sikhakolli, B. Vasantha and P. Tumuluru, "Analysis of Machine Learning based Feature Selection Method for Sentiment Analysis," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), pp. 700-705, 2021.

[12] I. Pujadayanti, M. Fauzi, S.Y. Arum, "Prediksi Rating Otomatis pada Ulasan Produk Kecantikan dengan Metode Naïve Bayes dan N-gram," Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, vol. 2, pp. 4421-4427, 2018.

[13] D. Ramayanti and U. Salamah, "Text Classification on Dataset of Marine and Fisheries Sciences Domain Using Random Forest Classifier," International Journal of Computer Techniques, vol. 5, pp. 1-7, 2018.

[14] A.R. Isnain, N.S. Marga, and D. Alita, "Sentiment Analysis Of Government Policy On Corona Case Using Naive Bayes Algorithm," Indonesian Journal of Computing and Cybernetics Systems, vol 15, pp. 55-64, 2021.

[15] M. Sadikin and F. Alfiandi, "Comparative Study of Classification Method on Customer Candidate Data to Predict its Potential Risk," International Journal of Electrical and Computer Engineering (IJECE), vol. 8, pp. 4763-4771, 2018.

[16] S. Khomsah, A.F. Hidayatullah, A.S. Aribowo, "Comparison of the Effects of Feature Selection and Tree-Based Ensemble Machine Learning for Sentiment Analysis on Indonesian YouTube Comments," Proceedings of the 1st International Conference on Electronics, Biomedical Engineering, and Health Informatics, pp. 161-172, 2021.

[17] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The Impact of Features Extraction on the Sentiment Analysis," *Procedia Computer Science,* vol. 152, pp. 341-348, 2019.

[18] P. Arya and A. Bhagat, "Performance Evaluation of Several Machine Learning Classification Algorithms with Combined Feature Selection Methods for Sentiment Analysis 704," International Journal of Innovative Technology and Exploring Engineering (IJITEE), pp. 703-710, 2019.

[19] D. Aggarwal, V. Bali, A. Agarwal, K. Poswal, M. Gupta, A. Gupta, "Sentiment Analysis of Tweets Using Supervised Machine Learning Techniques Based on Term Frequency," Journal of Information Technology Management, pp. 119-141, 2021.

[20] Aribowo, A. Sasmito et al. "Cross-domain sentiment analysis model on Indonesian YouTube comment," International Journal of Advances in Intelligent Informatics, pp. 12-25, 2021.

[21] Haryalesmana, D. (2017, 24 Maret). List of Opinion Words (positive/negative) in Bahasa Indonesia for Sentiment Analysis. Available: https://github.com/masdevid/ID-OpinionWords. (diakses tanggal 30 Juli 2021).

KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal yang telah di lampirkan sebelumnya dengan judul "Optimization of Various Naive Bayes and Decision Tree Experiments on E-Commerce Application Reviews". Kertas kerja ini berisi keseluruhan material hasil penelitan Tugas Akhir yang tidak dimuat atau disertakan dalam artikel jurnal. Di dalam kertas kerja ini menjelaskan penelitian secara jelas dan rinci mulai dari literatur review, dataset yang digunakan, tahapan eksperimen, source code, dan hasil pengolahan serta eksperimen secara keseluruhan akan lebih jelas di lampirkan. Di dalam kertas kerja disajikan:

- Bagian 1: Literatur Review
  Membahas mengenai literatur review yang berisi artikel jurnal dari penelitian sebelumnya dan teori yang menjadi dasar atau landasan dalam penelitian ini.
- Bagian 2: Dataset
  Membahas mengenai dataset yang digunakan dalam penelitian ini, meliputi: Cara perolehan data, sumber data, pelabelan data, dan sampel dari dataset.
- Bagian 3: Tahapan Penelitian
  Menjelaskan mengenai tahapan penelitian yang disajikan dalam bentuk alur diagram dengan penjelasan dari setiap tahapan dan skenario eksperimen yang dilakukan.
- Bagian 4: Source Code
- Bagian 5: Hasil semua eksperimen
  Menjelaskan hasil keseluruhan dari eksperimen yang telah dilakukan dengan penjelasan yang lebih lengkap.