IN
REVIEW

UNIVERSITAS
MERCU BUANA

Sentiment Classification Using Support Vector Machine on Indonesian Text Reviews of Mobile Application

TUGAS AKHIR

Hermawan
41515110157

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2020

Sentiment Classification Using Support Vector Machine on Indonesian Text Reviews of Mobile Application

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

Hermawan
41515110157

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2020

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM	:	41515110157
Nama	:	Hermawan
Judul Tugas Akhir	:	Sentiment Classification Using Support Vector Machine on Indonesian Text Reviews of Mobile Application

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 Februari 2020

Hermawan

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

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

Nama Mahasiswa	:	Hermawan
NIM	:	41515110157
Judul Tugas Akhir	:	Sentiment Classification Using Support Vector Machine on Indonesian Text Reviews of Mobile Application

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 Februari 2020

Hermawan

SURAT PERNYATAAN LUARAN TUGAS AKHIR

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

Nama Mahasiswa	:	Hermawan
NIM	:	41515110157
Judul Tugas Akhir	:	Sentiment Classification Using Support Vector Machine on Indonesian Text Reviews of Mobile Application

Menyatakan bahwa Luaran Tugas Akhir saya adalah sebagai berikut :

<table>
<tr><th>No</th><th>Luaran</th><th colspan="4">Jenis</th><th colspan="2">Status</th></tr>
<tr><td rowspan="6">1</td><td rowspan="4">Publikasi Ilmiah</td><td colspan="2">Jurnal Nasional Tidak Terakreditasi</td><td></td><td rowspan="2">Diajukan</td><td rowspan="2">✓</td></tr>
<tr><td colspan="2">Jurnal Nasional Terakreditasi</td><td></td></tr>
<tr><td colspan="2">Jurnal International Tidak Bereputasi</td><td></td><td rowspan="2">Diterima</td><td rowspan="2"></td></tr>
<tr><td colspan="2">Jurnal International Bereputasi</td><td>✓</td></tr>
<tr><td rowspan="2">Disubmit/dipublikasikan di :</td><td>Nama Jurnal</td><td colspan="4">: IAES International Journal of Artificial Intelligence (IJ-AI)</td></tr>
<tr><td>ISSN</td><td colspan="4">: ISSN/e-ISSN 2089-4872/2252-8938</td></tr>
<tr><td rowspan="7">2</td><td rowspan="7">Kertas Kerja, Merupakan material haasil penelitian sebagai kelengkapan Artikel Jurnal. Terdiri dari (minimal 4)</td><td colspan="5">Literatur Review</td><td>[✓]</td></tr>
<tr><td colspan="5">Hasil analisa & perancangan aplikasi</td><td>[ ]</td></tr>
<tr><td colspan="5">Source code</td><td>[✓]</td></tr>
<tr><td colspan="5">Data set</td><td>[✓]</td></tr>
<tr><td colspan="5">Tahapan eksperimen</td><td>[✓]</td></tr>
<tr><td colspan="5">Hasil eksperimen seluruhnya</td><td>[✓]</td></tr>
<tr><td colspan="5">....</td><td></td></tr>
<tr><td rowspan="4">3</td><td rowspan="4">HAKI
Disubmit / Terdaftar</td><td colspan="3">HKI</td><td></td><td>Diajukan</td><td></td></tr>
<tr><td colspan="3">Paten</td><td></td><td>Tercatat</td><td></td></tr>
<tr><td>No & Tanggal Permohonan</td><td colspan="4">:</td></tr>
<tr><td>No & Tanggal Pencatatan</td><td colspan="4">:</td></tr>
</table>

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 12 Februari 2020

Hermawan

LEMBAR PERSETUJUAN

Nama Mahasiswa	:	Hermawan
NIM	:	41515110157
Judul Tugas Akhir	:	Sentiment Classification on Mobile Application Reviews Using Support Vector Machine

Tugas Akhir ini telah diperiksa dan disetujui

Jakarta, 14 Desember 2019

Menyetujui,

(Drs.Achmad Kodar, MT)
Dosen Pembimbing

LEMBAR PERSETUJUAN PENGUJI

NIM	:	41515110157
Nama	:	Hermawan
Judul Tugas Akhir	:	Sentiment Classification Using Support Vector Machine on Indonesian Text Reviews of Mobile Application

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, 12 Februari 2020

(Dr. Ir. Eliyani)
Ketua Penguji

(Harni Kusniyati, ST, M.Kom)
Anggota Penguji 1

(Afiyati, S.Si, MT)
Anggota Penguji 2

LEMBAR PENGESAHAN

NIM	:	41515110157
Nama	:	Hermawan
Judul Tugas Akhir	:	Sentiment Classification Using Support Vector Machine on Indonesian Text Reviews of Mobile Application

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, 12 Februari 2020

Menyetujui,

(Drs.Achmad Kodar, MT)
Dosen Pembimbing

Mengetahui,

(Diky Firdaus, S.Kom, MM)	(Desi Ramayanti, S.Kom, MT)
Koord. Tugas Akhir Teknik Informatika	Ka. Prodi Teknik Informatika

ABSTRAK

Nama	:	Hermawan
NIM	:	41515110157
Pembimbing TA	:	Drs. Achmad Kodar, MT
Judul	:	Klasifikasi Sentimen Menggunakan Support Vector Machine Pada Ulasan Aplikasi Mobile Berbahasa Indonesia

Saat ini, ulasan aplikasi pada toko penyedia aplikasi mobile menjadi sumber daya informasi yang sangat bermanfaat bagi para developer. Namun banyaknya data yang tidak terstruktur pada ulasan, developer harus menerjemahkan ulasan tersebut agar dapat mengetahui bagian mana yang akan dilakukan perbaikan pada aplikasi. Studi penelitian tentang klasifikasi sentimen sudah banyak dilakukan dengan sebagian besar penelitian diterapkan pada analisa menggunakan bahasa inggris. Penelitian ini bertujuan untuk mengimplementasikan algoritma Support Vector Machine pada klasifikasi sentimen ulasan aplikasi berbahasa indonesia menggunakan bahasa pemrograman python dan library scikit-learn. Himpunan data ulasan diambil secara manual sebanyak 1503 kalimat data dari google play. Kemudian dataset dilabeli setiap kalimatnya dengan sentimen positif dan negatif. Sebagai hasilnya percobaan dengan klasifikasi SVM dengan fitur TF-IDF dengan perbandingan rasio data training 80% dan data testing 20% mencapai akurasi 0,923. Dari hasil akurasi tersebut data model dilakukan sebuah validasi dengan memberikan data kalimat ulasan baru, hasilnya mesin sudah bisa memprediksi hasil sentimen yang diharapkan

Kata kunci:
Klasifikasi teks, Machine learning, Sentimen analisis, Support Vector Machine, Ulasan aplikasi

ABSTRACT

Name	:	Hermawan
Student Number	:	41515110157
Counsellor	:	Drs. Achmad Kodar, MT
Title	:	Sentiment Classification Using Support Vector Machine on Indonesian Text Reviews of Mobile Application

Currently, reviews on mobile application stores become very useful information resources for developers. However, due to the large amount of unstructured data on the reviews, the developer must interpret the review in order to know which parts will be improved on the application. Research studies on sentiment classification have been conducted with most of which are applied to analysis using English. This study aimed to implement the Support Vector Machine algorithm in the sentiment classification of reviews in Indonesian language applications using a programming language of Python and Scikit-learn Library. The collection of review data was taken manually which were 1503 data sentences from Google Play. Then the dataset was labelled for its every sentence with positive and negative sentiments. As the result, the sentiment classification using SVM and TD-IDF features showed the best performance having the accuracy result of 0.923, with the ratio of 80% training data and 20% testing data. The implementation of model data was done by providing new review sentence data. In consequence, the machine is already able to predict the expected sentiment.

Key words:
Application review, Machine learning, Sentiment analysis, Support vector machine, Text classification.

KATA PENGANTAR

Puji syukur kita panjatkan kepada Allah Subhanahu Wa Ta'ala yang telah melimpahkan rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan tugas akhir yang berjudul "Sentiment Classification Using Support Vector Machine on Indonesian Text Reviews of Mobile Application" tepat pada waktunya. Tugas akhir ini disusun untuk memenuhi salah satu syarat memperoleh gelar sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.

Penulis menyadari bahwa tanpa bantuan dan bimbingan yang melibatkan banyak pihak, penelitian ini tidak akan terlaksana dengan baik. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Bapak Drs.Achmad Kodar, MT, selaku dosen pembimbing tugas akhir yang telah meluangkan waktunya untuk memberikan bimbingan serta arahan dalam penyusunan tugas akhir ini hingga selesai.
2. Ibu Desi Ramayanti, S.Kom., MT, selaku Kepala Prodi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana.
3. Bapak Diky Firdaus, S.Kom., MM, selaku Koordinator Tugas Akhir Program Studi Teknik Informatika, Fakultas Ilmu Komputer Universitas Mercu Buana tahun ajaran 2019/2020.
4. Kedua Orang tua yang senantiasa memberikan doa dan dukungan kepada penulis.
5. Sulistiningsi, S.I.Kom, selaku *reviewer* pada penulisan jurnal ilmiah.
6. Teman-teman yang selalu saling menyemangati dan berkerja sama selama pelaksanaan tugas akhir.
7. Semua pihak yang telah banyak membantu dalam penyusunan tugas akhir ini yang tidak dapat penulis sebutkan satu persatu

Akhir kata, penulis berharap tugas akhir ini dapat bermanfaat bagi pembaca guna menambah pengetahuan dan wawasan.

Jakarta, 15 Februari 2020

Hermawan

DAFTAR ISI

NASKAH JURNAL

Sentiment Classification on Mobile Application Reviews Using Support Vector Machine

Hermawan[1], Achmad Kodar[2]
[1,2] Department of Informatics Engineering, Universitas Mercu Buana, Indonesia

Article Info

Article history:

Keywords:

Application review
Machine learning
Sentiment analysis
Support vector machine

ABSTRACT

Currently, reviews on mobile application stores become very useful information resources for developers. However, due to the large amount of unstructured data on the reviews, the developer must interpret the review in order to know which parts will be improved on the application. Research studies on sentiment classification have been conducted with most of which are applied to analysis using English. This study aimed to implement the Support Vector Machine algorithm in the sentiment classification of reviews in Indonesian language applications using a programming language of *Python* and *Scikit-learn* Library. The collection of review data was taken manually which were 1503 data sentences from Google Play. Then the dataset was labelled for its every sentence with positive and negative sentiments. As the result, the sentiment classification using SVM and TD-IDF features showed the best performance having the accuracy result of 0.923, with the ratio of 80% training data and 20% testing data. The implementation of model data was done by providing new review sentence data. In consequence, the machine is already able to predict the expected sentiment.

Corresponding Author:

Hermawan,
Department of Informatics Engineering,
Universitas Mercu Buana,
Jakarta, Indonesia.
Email: 41515110157@student.mercubuana.ac.id

1. INTRODUCTION

The number of downloads on mobile applications has increased along with the increase in smartphone users. Based on the data released by *Appannie* [1], the number of wordwide downloads in 2018 increased by 35% from 2016 with total downloads reaching 194 billion. This is a developer's challenge to continue to compete in showing the best applications. Thus, there will be a variety of new and creative applications rising up in the mobile application provider stores.

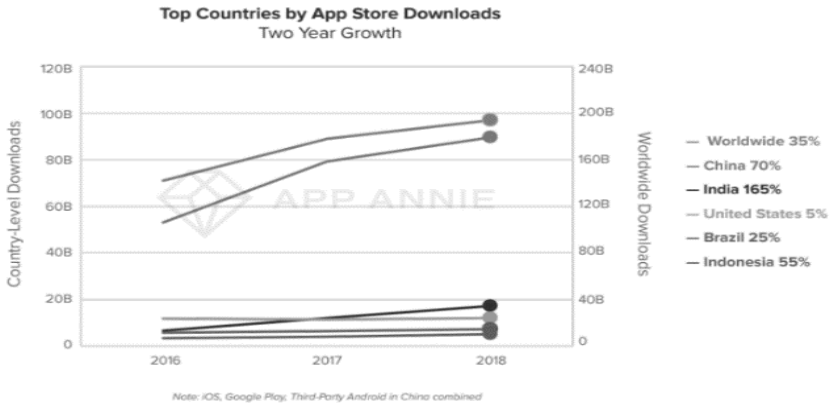


Figure 1. Improvement data by Appannie [1]

Google Play as one of the mobile application providers, offers a review feature for app developers as useful information for developing higher quality applications. This feature takes the form of feedback from the user to express their opinions and sentiments towards the application downloaded, in the form of ratings and text reviews [2]. Although developers can use ratings to assess the level of user satisfaction towards the application, they must still interpret the text review to find out the advantages and disadvantages of the application.

Research studies on sentiment classification on reviews have been widely carried out [3-5]. However, most of these studies use English as a dataset. Sentiment classification can use one type of artificial intelligence approach called machine learning. The role of machine learning is needed for doing fast analysis and automation of big data using a computer algorithm in it [6]. The results of the classification are expected to be used as a reference in making fast and valid decisions on business processes in the future [7,8].

One of the text classification algorithms that can be applied for sentiment analysis is Support Vector Machine. The SVM algorithm has been used extensively in previous studies. One of which was [9] who classified the Indonesian-language complaint text data on the tweets data of the Ministry of Maritime Affairs and Fisheries of the Republic of Indonesia *(@ kkpgoid)*. There was also a study by [10], classify grain using image processing techniques with Support Vector Machine. This algorithm excels in its high level and accuracy for modelling in a sentiment classification [11-13].

Depart from several previous studies, this study will use the Support Vector Machine Algorithm with Indonesian language text review data taken manually from various applications available on Google Play. The review classification process will be divided into two classes of sentiment polarity, namely positive sentiment and negative sentiment classes. The use of SVM algorithm in this study aimed to find out how high the level of accuracy of sentiment classification and the use of model data results which can be exerted as information in the future development of applications and business processes.

2. RESEARCH METHOD

The research methodology will be completed in six stages, starting from data collection, data labelling, data pre-processing, algorithm classification, feature selection, and evaluation, as illustrated in Figure 2 below.

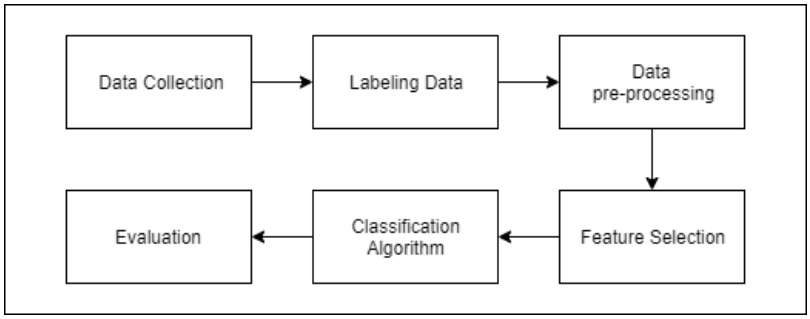


Figure 2. Research Method

2.1. Data Collection

The initial stage of collecting data review was done manually taken from various applications on the Google Play website. Data were collected based on satisfaction and dissatisfaction ratings from users. Reviews or comments taken were only in Indonesian text with a final data reaching 1503 review sentences. The example of a comment sentence taken was as follows.

Table 1. Sample reviews

No	Comment
1	Aplikasinya sangat membantu dalam kehidupan saya di kerjaaan. Terima kasih
2	UI nya terlalu ribet, orang awam akan bingung pakai aplikasi ini. Tolong perbaikannya
3	Aplikasi sampah!!!! Banyak iklan koneksi lemot buat transfer data. Gak jelas! Pake thatering Bluetooth semua koneksi dipakai tapi file tidak terkirim dengan baik!!!!!

2.2. Data labelling

In the case of classification, labelling correctly was the most important thing in order for the data to be collected or classified according to the class. At this stage, the data collected were classified into two classes of sentiments, namely positive and negative sentiments. Each review sentence was labelled sentiment according to the meaning of the review, satisfaction or dissatisfaction. Figure 3 explains the comparison of the percentage of sentiments with the acquisition of 752 positive sentences and 751 negative sentences.

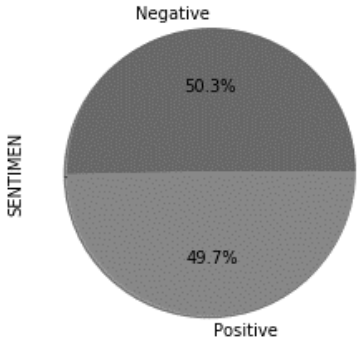


Figure 3. Presentage of sentiments

2.3. Data pre-processing

At this stage, the data labelled with sentiment underwent the data cleaning process, which is a pre-processing stage to select data to be more structured. The pre-processing stage involved one of the *Python* libraries that was the literary *Sastrawi* library. The pre-processing stages included:

a. Clean number

Cleaning the sentence data was done by removing all numeric numbers (0 - 9). The results of this process will only leave characters starting from the alphabet with the letter *a* to the letter z.

b. Deleting the Punctuation

Deleting symbols was done to delete special characters in comments such as punctuation (commas (,), periods (.), question marks (?), exclamation points (!), and etc.), and other characters ($, %, *, and etc).

c. Case folding

Here, all uppercase letters in a comment were changed to lowercase characters. Only alphabets with the letters 'a' to 'z' are accepted. Characters other than letters are omitted and are considered delimiter.

d. Clean One Character

In this stage, single characters having no meaning were deleted. One letter in question is for example y, g, and so forth, even though the character, say "*y"* character means "*ya"* and "*g"* means "*gak"*. So for the extraction process, the data is a word that is not easily declared because it has no clear meaning.

e. Stopword removal

Stopword removal is intended to find out whether a word has entered into a stopword or not. This process will remove unimportant words in the dataset such as words such as *yang*, *tetapi*, *atau*, *ke*, *di*, *dengan*, and others.

f. Stemming

In this stage, words were changed into root words by eliminating all word affixes including word prefixes, word insertions, word suffixes and/or eliminating word prefixes and suffixes in derived words. As the example, the word "*dibantu*", "*membantu*", "*terbantukan*", will be changed into the root word, which is "*bantu*".

2.4. Algorithm classification

Support Vector Machine is a supervised classification technique that can be used in statistics and machine learning [14]. The working principle of the SVM method seeks to find the optimal *hyperplane* of a dataset and maximize the margin between the two classes [15]. *Optimal* meant here is the boundary line which can separate 2 groups with the farthest distance between the outermost points in each group and the boundary line itself [16].

2.5.Feature selection

At this stage, TF-IDF feature was used as the experiment, whether using the feature selection can improve accuracy results or not. TF-IDF method is a method for calculating the weight of each of most frequently appearing or used words. This method will calculate the value of Term Frequency (TF) and Inverse Document Frequency (IDF) on each review word in the dataset.

2.6.Evaluation

The last stage was evaluating the data accuracy results with a prediction percentage and the actual conditions of the classification results using the SVM algorithm by applying *classification_report, confusion_matrix,* and *accuracy_score.* Table 2 below illustrates the concept of confusion matrix:

Table 2. Confusion Matrix

		Predict	
		Negative	Positive
Actual	Negative	***TP***	***FP***
	Positive	***TN***	***FN***

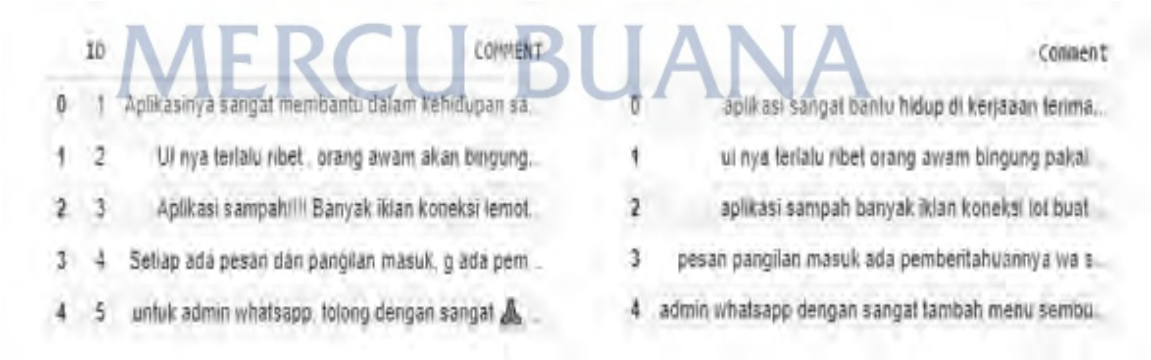
Based on the concept of confusion matrix, performance calculation indicators are TP, FP, TN, and FN. The explanation is presented in table 3 below.

Table 3. Name and definition of confusion matrix

Name	**Definition**
TP	The number of positive data considered true
FP	The number of positive data considered false
TN	The number of negative data considered true
FN	The number of negative data considered false

3. RESULTS AND DISCUSSION

Data cleaning process at the pre-processing stage was completed by using *Sastrawi* library. The time needed to extract data from 1503 sentences to be clean was 3 minutes 11 seconds. The top five review sentences resulting from cleaning data are illustrated in Figure 4 below as a comparison of word changes.



	ID	COMMENT
0	1	Aplikasinya sangat membantu dalam kehidupan sa...
1	2	UI nya terlalu ribet , orang awam akan bingung...
2	3	Aplikasi sampah!!!! Banyak iklan koneksi lemot...
3	4	Setiap ada pesan dan panggilan masuk, g ada pem...
4	5	untuk admin whatsapp, tolong dengan sangat 🙏...

	Comment
0	aplikasi sangat bantu hidup di kerjaan terima...
1	ui nya terlalu ribet orang awam bingung pakai...
2	aplikasi sampah banyak iklan koneksi lot buat...
3	pesan pangilan masuk ada pemberitahuannya wa s...
4	admin whatsapp dengan sangat tambah menu sembu...

(a) (b)

Figure 4. Result of pre-processing
(a) before, (b) after

The cleaning data results from the pre-processing stage was used for the classification process of sentiment reviews using the SVM classifier in the *Scikit-learn* library. The data validation process was completed by separating data

automatically using Split Data. The data set was separated into two parts; the first was the training data used for the model and the second was the testing data used to measure accuracy performance.

The experiment of SVM classification was carried out by tuning parameters in order to find out how much influence the parameter tuning had on the accuracy value. These parameters included the use of feature selection and the comparison of ratios between training data and testing data. The experiment of ratio comparison was done five times with a split data comparison scheme from the ratio of 50:50 to 90:10. Each ratio will be given two different test scenarios. The first scenario used the TF-IDF feature, while the second did not. The results of SVM classification from experiments with tuning parameters are shown in table 4 below.

Table 4. Experiment result

Dataset	Classifier	Training v/s Testing	Feature	Accuracy
Riview Application	Support Vector Machine	90:10	TF-IDF	0.900
			NO	0.894
		80:20	TF-IDF	0.923
			NO	0.893
		70:30	TF-IDF	0.917
			NO	0.880
		60:40	TF-IDF	0.905
			NO	0.872
		50:50	TF-IDF	0.902
			NO	0.871

Table 4 explains the experiment results of the SVM classification with tuning parameter. The final results proved that the greatest accuracy was found in the 90% ratio of training data under the scenario of without TF-IDF feature. The more increasing the number of training data, the higher the accuracy value. This is because when more training data is owned, the machine will have more knowledge as learning. The second scenario using the TF-IDF feature made a difference, in which the best accuracy was owned by 80% ratio with the accuracy value of 0.923. Table 5 will explain the evaluation of the best accuracy data from that ratio with detailed accuracy using the confusion matrix, classification report, and accuracy score.

Table 5. confusion matrix, classification report, and accuracy score

N = 1503		Predicted Class		precision	recall	f1-score
		Negative	Positive			
Actual Class	Negative	140	15	0.90	0.95	0.92
	Positive	8	138	0.95	0.90	0.92
Accuracy						0.92

Based on experiments, the information that can be obtained from table 5 is the number of accuracy score achieved was 0.923. In this model, prediction of new review data can be already implemented to predict sentiment. The factor affecting the accuracy results was the consistency of sentiment labelling in each

review sentence. The number of prediction errors is explained in the confusion matrix table above, with the following information:

- 140 is TRUE predictive value from negative value
- 8 is FALSE predictive value from positive value
- 15 is FALSE predictive value from negative value
- 138 is TRUE predictive value from positive value

4. CONCLUSION

Based on the research results that has been conducted to determine the sentiments of the review of mobile applications in Indonesian language using the SVM classifier in the *sklearn* library, some conclusions can be drawn as follows:

- The addition of TF-IDF feature on SVM Classifier with various training data and testing data ratios can affect the accuracy results. The accuracy will be higher than without using TF-IDF.
- The best ratio with the highest accuracy score was found in the comparison between 80% training data and 20% testing data. The results reached an accuracy of 0.923 with precision and recall values respectively 0.90 and 0.95 for the negative sentiment dataset and 0.95 and 0.90 for the positive sentiment dataset.
- With these accuracy results, the results of the data model can already predict the sentiment of a new review data.

ACKNOWLEDGEMENTS

The author would like to thank everyone who contributes directly or undirectly to this research. Especially Ms. Sulistiningsi who has been the reviewer of this article.

REFERENCES'

[1] Appannie, "The State Of Mobile 2019," 2019. [Online]. Available: https://s3.amazonaws.com/files.appannie.com/reports/1901_State_of_Mobile_Main_EN.pdf. [Accessed: 19-Nov-2019].

[2] Google, "Write a review on Google Play - Computer - Google Play Help." [Online]. Available: https://support.google.com/googleplay/answer/4346705. [Accessed: 25-Nov-2019].

[3] A. Salinca, "Business Reviews Classification Using Sentiment Analysis," *Proc. - 17th Int. Symp. Symb. Numer. Algorithms Sci. Comput. SYNASC 2015*, no. April, pp. 247–250, 2016.

[4] K. Varshney, "Sentiment Analysis of Application Reviews on Play Store," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 2, no. 2, pp. 510–513, 2017.

[5] J. Ding, H. Sun, X. Wang, and X. Liu, "Entity-level sentiment analysis of issue comments," *Proc. - Int. Conf. Softw. Eng.*, no. June, pp. 7–13, 2018.

[6] A. A. Hania and Y. T. Indonesia, "Mengenal Artificial Intelligence , Machine Learning , Neural Network , dan Deep Learning," *yayasan cahaya Islam. J. Teknol. Indones.*, no. Juni, 2017.

[7] T. C. Dewi, W. Maharani, and A. Herdiani, "analisis klasifikasi sentimen level aspek menggunakan pendekatan pembelajaran supervised aspect level sentiment classification based on supervised," *e-Proceeding Eng.*, vol. 3, no. 2, pp. 3654–3660, 2016.

[8] I. Zulfa and E. Winarko, "Sentimen Analisis Tweet Berbahasa Indonesia dengan Deep Belief Network," *Indones. J. Comput. Cybern. Syst.*, vol. 11, no. 2, pp. 187–198, 2017.

[9] D. Ramayanti and U. Salamah, "Complaint Classification Using Support Vector Machine for Indonesian Text Dataset," *Int. J. Sci. Res. Comput. Sci.*, vol. 3, no. 7, pp. 179–185, 2018.

[10] S. Ibrahim, N. A. Zulkifli, N. Sabri, A. A. Shari, and M. R. M. Noordin, "Rice grain classification using multi-class support vector machine (SVM)," *IAES Int. J. Artif. Intell.*, vol. 8, no. 3, pp. 215–220, 2019.

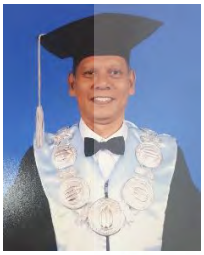
[11] S. Nurwahyuni, U. Bina, and S. Informatika, "Online KRL access menggunakan metode Naive Bayes," *Swabumi*, vol. 7, no. 1, pp. 31–38, 2019.

[12] A. Charim, S. Basuki, D. R. Akbi, and U. M. Malang, "Detect Malware in Portable Document Format Files (PDF) Using Support Vector Machine and Random Decision Forest," *J. Online Inform.*, vol. 3, no. 2, pp. 99–102, 2019.
[13] V. D. Kaur, "Sentimental Analysis of Book Reviews using Unsupervised Semantic Orientation and Supervised Machine Learning Approaches," *Second Int. Conf. Green Comput. Internet Things*, pp. 519–524, 2018.
[14] V. Ayumi and M. I. Fanany, "A comparison of SVM and RVM for human action recognition," *Internetworking Indones. J.*, vol. 8, no. 1, pp. 29–33, 2016.
[15] K. Waktu, K. Mahasiswa, A. Pratama, R. C. Wihandika, and D. E. Ratnawati, "Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa," *J. Pengemb. Teknol. Inf. dan ilmu Komput.*, vol. 2, pp. 1704–1708, 2018.
[16] "Machine Learning: Support Vector Machines - Artificial intelligence and data science." [Online]. Available: https://www.megabagus.id/machine-learning-support-vector-machines/. [Accessed: 09-Dec-2019].

BIOGRAPHIES OF AUTHORS

Hermawan is a final year undergraduate student in Informatics Engineering, Faculty of Computer Science, Universitas Mercu Buana, Indonesia under the supervision of Achmad Kodar. He was born in jakarta on October 8th, 1992. His research interests are related to big data, data mining, and machine learning.

E-mail address: 41515110157@student.mercubuana.ac.id.



Achmad Kodar received his bachelor's degree in Mathematics at Padjadjaran University. He has two master degrees in Industrial Engineering from the University of Indonesia and in Computer Engineering from Budi Luhur University. He has been a teaching staff / lecturer in the Faculty of Computer Science, Mercu Buana University since 1993. His teaching history is he frequently teaches cryptographic and calculus materials. Besides being a teaching lecturer, he is also the supervisor of his students' final assignments.

E-mail address: achmad.kodar@mercubuana.ac.id.

KERTAS KERJA

Ringkasan

Saat ini, ulasan aplikasi pada toko penyedia aplikasi mobile menjadi sumber daya informasi yang sangat bermanfaat bagi para *developer*. Namun banyaknya data yang tidak terstruktur pada ulasan, *developer* harus menerjemahkan ulasan tersebut agar dapat mengetahui bagian mana yang akan dilakukan perbaikan pada aplikasi. Studi penelitian tentang klasifikasi sentimen sudah banyak dilakukan dengan sebagian besar penelitian diterapkan pada analisa menggunakan bahasa inggris. Penelitian ini bertujuan untuk mengimplementasikan algoritma *Support Vector Machine* pada klasifikasi sentimen ulasan aplikasi berbahasa indonesia menggunakan bahasa pemrograman *python* dan *library scikit-learn*. Himpunan data ulasan diambil secara manual sebanyak 1503 kalimat data dari *google play*. Kemudian dataset dilabeli setiap kalimatnya dengan sentimen positif dan negatif. Sebagai hasilnya percobaan dengan klasifikasi *SVM* dengan fitur *TF-IDF* dengan perbandingan rasio data training 80% dan data testing 20% mencapai akurasi 0,923. Dari hasil akurasi tersebut data model dilakukan sebuah validasi dengan memberikan data kalimat ulasan baru, hasilnya mesin sudah bisa memprediksi hasil sentimen yang diharapkan.