

# Recommendation for Classification of News Categories Using Support Vector Machine Algorithm with SVD

Nofenky<sup>1</sup>, Dionisia Bhisetya Rarasati<sup>2</sup>

Fakultas Teknologi dan Design, Universitas Bunda Mulia, Jakarta, Indonesia

<sup>1</sup>nofenky@gmail.com, <sup>2</sup>rarasati.ubm@gmail.com

Accepted 30 November 2020

Approved 10 August 2021

**Abstract**— online news is a digital information media currently has a very easy and flexible updating process. The News Document grouping process is implemented in several stages, including Text Mining which includes Text Pre-processing which includes Tokenizing, Stopword removal, Stemming, Word Merging, TF-IDF and Confusion Matrix. Of the several techniques in Text Mining, the most frequently used for News Document classification is Support Vector Machine (SVM) Algorithm. SVM has many advantages of being able to identify separate hyperplane that maximizes the margin between two or more different classes. The selection of the features in SVM Algorithm significantly affects the classification accuracy results. Therefore, in this study a combination of the feature selection methods is used, namely Singular Value Decomposition in order to increase accuracy and reduce the Classifier Time Support Vector Machine. This research resulted in text classification in the form of categories Entertainment, Health, Politics and Technology. Based on the Support Vector Machines Algorithm, an accuracy rate of 81% was obtained with 360 Data Training and 120 Data Testing, after adding the Singular Value Decomposition feature with a K-Rank value of 50%, a significant increase in accuracy was obtained with an accuracy value of 94% and The time of Algorithm process is faster.

**Index Terms**— News Classification, Support Vector Machine, Singular Value Decomposition, Text Mining.

## I. INTRODUCTION

Information has become one of the necessities in everyday human life. Information can be interpreted as knowledge from a learning, experience or instruction. In some cases knowledge about events or situations that have been collected or received through the communication process, gathering news from certain events.[1] In the modern era and the sophistication of technology, speed and accuracy in obtaining news or information are needed by the community. Newspapers are the people's choice in obtaining fast and accurate news and information.

In general, the news that is conveyed in news portals consists of various categories such as Political News, Technology News, Entertainment News, Health News and others. For example on the Indonesian news

website [www.kompas.com](http://www.kompas.com), [www.waspada.com](http://www.waspada.com), [www.vivanews.com](http://www.vivanews.com), etc. [2]

However, in distributing news into certain categories, for now it is still done manually, where the news is separated by the author, meaning that in uploading news the writers must first know what the content of the news they will upload will be and then it will be included in the right and correct category. Therefore, it is necessary to have a system that can classify news categories automatically according to the existing news categories based on the title and content of the news so that this system can assist news uploaders in uploading news.

The SVM algorithm can be explained simply as an attempt to find the best hyperplane that can function as a separator between 2 classes in the input space. The SVM method is rooted in statistical learning theories where this method is very promising in providing better results than some other similar methods. [3]

To be able to produce higher accuracy than a Support Vector Machine Algorithm, an experiment will be carried out by adding the Singular Value Decomposition (SVD) Method, SVD is a matrix decomposition technique to facilitate data processing, because the Singular Value Decomposition (SVD) Algorithm has advantages over process time efficiency Algorithm on a large-scale Dataset and Singular Value Decomposition (SVD) Algorithm were also chosen because they have the ability to perform the decomposition process on a term-document matrix, so that a matrix that still stores important information with far-reaching smaller dimensions can be obtained.

The purpose of this research is to build a system that can automatically classify news categories into a news category into an actual category, to find out the difference in the final results of News Classification with the Support Vector Machine Algorithm only and with the addition of the Singular Value Decomposition Method, and also to make it easier for news uploaders to upload news with an automatic news category categorization system.

## II. LITERATURE REVIEW

### A. Text Mining

Text mining is a technique in computer science that can be used to solve a large number of information problems by combining techniques from Data Mining, Machine Learning, Natural Language Processing, Information Retrieval and Knowledge Management. [4]. Text Mining seeks to extract useful information from data sources through an identification and exploration of interesting patterns. In Text Mining the data source can be obtained from a collection of documents, this means that the data can be in the form of newspapers, magazines, articles, letters, or research reports such as journals, or a thesis.

### B. Stages of Text Mining

An explanation of each of the Text Preprocessing processes is as follows:[5]

#### 1). Tokenizing.

Tokenizing Text is unstructured data that must be changed first to make it structured before further analysis. The text in the email entered into the application is stored in a 1-dimensional array. The words in a sentence are divided based on the sentence and then the words will be divided again based on spaces.

#### 2). Stopword removal.

After doing the Tokenizing Text process that the word is not tied to other words. As a result of this separation, there are some words that have no relevant meaning at all in determining the characteristics of a tokenized document, such as the words "this, that, and, or" and many more words. kind. Words that have no relevant meaning are called stopwords.

#### 3). Stemming.

Stemming is a process of mapping and parsing in the Variant form of a word into its basic word form. In Indonesian documents the Stemming process is very necessary before entering the Text Mining process because Indonesian has Prefixes, Suffixes, Infixes and Confixes which make a basic word can be changed into many forms and as a result of making word searches difficult. The following are examples and meanings of affixes in Indonesian.

- a. *Suffixes* (Akhiran) is an affix that is usually added to the end of a word, such as "-an, -kan, and '-i".
- b. *Prefixes* (Awalan) is an affix that is usually added to the beginning of a root word or the basic form of a word, such as "-per, -mem"
- c. *Confixes* (Sifiks and Prefiks) a single affix occurs from two separate elements, such as "ke-....-an"

#### 4). Word weighting.

The TF-IDF method is a method for calculating the weight of each word that is most commonly used in Information Retrieval. This method is also known to be efficient, easy to use and has high accuracy results.[6]

The data that has gone through the preprocessing stage is in the form of a numeric using this method TF-IDF. The Term Frequency Inverse Document Frequency (TF-IDF) method is a method commonly used to determine how far connected a word (term) is to a document by giving weight to each word. The TF-IDF method itself combines two concepts, namely the frequency of occurrence of a word in a document and the inverse frequency of a document containing these words.[7]

In calculating the weight with TF-IDF, what is first calculated is the TF value of a word with the weight of each word being 1. TF (Term Frequency) which states the number of words that appear in a document. DF (Document Frequency) states how many documents for TF calculation using the following formula: [8]

$$DF = \log \left( \frac{N}{DF(w)} \right) \quad (1)$$

$$TF - IDF(w, d) = TF(w, d) \times IDF(w) \quad (2)$$

$$IDF(Word) = \log \frac{td}{df} \quad (3)$$

Explanation:

- $TF-IDF(w,d)$ : the weight of one word in the entire document.
- $w$ : word.
- $d$ : document.
- $TF(w,d)$ : frequency of occurrence of a word ( $w$ ) in a document ( $d$ ).
- $IDF(w)$ : inverse  $DF$  from word ( $w$ ).
- $N$ : total number of documents.
- $DF(w)$ : the number of documents containing a word ( $w$ ).

$IDF(word)$  is an  $IDF$  value of each word to be searched, while  $TD$  is the total number of existing documents and  $DF$  is the number of occurrences of words in a document.

### C. Confusion Matrix

Confusion Matrix is the method used in the calculation of accuracy. In testing the accuracy and search results will be evaluated into the value of Recall, Precision and Accuracy. Where Precision is an evaluation of the ability of the system to find the most relevant ranking and can be defined as a percentage of documents that are retrieved and are truly relevant to the Query. Recall is an evaluation of the system's ability to find all relevant items from collections and can be defined as a percentage of documents relevant to the Query. Meanwhile, Accuracy is a comparison between

cases that will be correctly identified and the total number of existing cases.[6]

Table 1. Confusion Matrix.

Document	True Value	
	Relevant	Non Relevant
Retrieve d	True Positive (tp) Correct result	False Positive (fp) Unexpected result
Not Retrieve d	False Negative (fn) Missing Renault	True Negative (tn) Correct absence of Renault

Explanation:

- *TP (True Positive)* = Number of correctly labelled positive samples.
- *FP (False Positive)* = Number of negative samples incorrectly labelled as positive.
- *FN (False Negative)* = Number of positive samples incorrectly labelled as negative
- *TN (True Negative)* = Number of correctly labelled negative samples

formula as follows:

$$precision = \frac{tp}{(tp+fp)} \quad (4)$$

$$Recall = \frac{tp}{(tp+fn)} \quad (5)$$

$$Accuracy = \frac{(tp+tn)}{(tp+fp+tn+fn)} \quad (6)$$

#### D. Support Vector Machine

Support Vector Machine (SVM) is a technique that is relatively new when compared to other existing techniques, but SVM has a much better performance in various application fields such as Bioformatics, Handwriting Recognition, Text Classification and so on. [3]

SVM is a technique for making predictions, both in the case of classification and regression. SVM has the basic principle of linear classifier which means that linear classification cases can be separated, but SVM has been developed so that it can work on non-linear problems by including kernel concepts in a high-dimensional workspace. The kernel function is commonly used to map a lower initial dimension to a relatively higher dimension. Kinds of kernel functions include:[10]

Formula Support Vector Machine

$$f(xd) = \sum a_i y_i K(x_i, x_d) + b \quad (x_i s v) \quad n s \quad i=1(7)$$

Kernel Function

- Kernel linier  
 $K(x_i, x) = x_i T x \quad (1)$

- Polynomial  
 $K(x_i, x) = (\gamma \cdot x_i T x + r)^p, \gamma > 0 \quad (9)$

- Radial basis function  
 $K(x_i, x) = \exp(-\gamma |x_i - x|^2), \gamma > 0 \quad (10)$

- Sigmoid kernel  
 $K(x_i, x) = \tanh(\gamma \cdot x_i T + r) \quad (11)$

#### E. Singular Value Decomposition (SVD)

Singular Value Decomposition is one of the many techniques in processing a matrix derived from the science of linear algebra which was introduced by Beltrami in 1873. SVD is one of the stages in the process contained in the Latent Semantic Analysis (LSA) method. used to process a matrix in linear algebra which is used as a tool in mathematics and is commonly used to represent a matrix and is capable of performing various analyzes and matrix computations.[11] SVD is very useful in decomposing a matrix divided into 3 new matrices, including the U orthogonal matrix, the S diagonal matrix and the last one is the transpose matrix of the D orthogonal matrix or it can also be formulated as follows:

Formula Singular Value Decomposition.

$$A_{m \times n} = U_{m \times n} S_{m \times n} V_{n \times n}^t \quad (12)$$

Explanation:

- $m \times n$  is a matrix
- A Matrix size  $m \times n$
- U The singular vector of the matrix A and this vector is orthonormal
- S The diagonal of the vector that composes the singular value of the corresponding singular vector.
- VT The singular vector of matrix A is also orthonormal.

#### F. F1-Score

F1 or F-Measure is a harmonic mean of precision and recall or it can be abbreviated as f1-score which is a comparison of the mean or average of precision and recall that is weighted. The range of an f1 value is 0 to 1. Here's the equation:[11]

$$f1 = 2 \times \frac{P \times R}{P+R} \quad (13)$$

Explanation:

- F1 : F-Measure or F-Score
- P : Precision
- R : Recall

### III. RESEARCH METODOLOGY

In this chapter, it is explained about the stages in system design that will be made by the author, which includes Data, System Description, Analysis Model and Interface Design. Here's the explanation

**A. Development Method**

The method of developing a system used in this research is the Waterfall Method or Structured Approach. In general, the Waterfall method is a method that is often used to analyze systems. The essence of this method is the stages in working on a system that are carried out sequentially, the Waterfall method consists of several stages of activities, including:

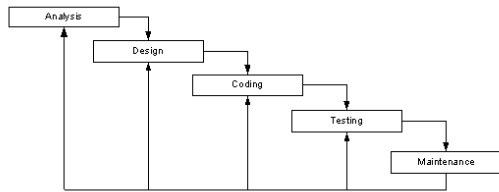


Figure 1. Development Method waterfall.

**B. Diagram Block**

Below is a system process using Block Diagrams:

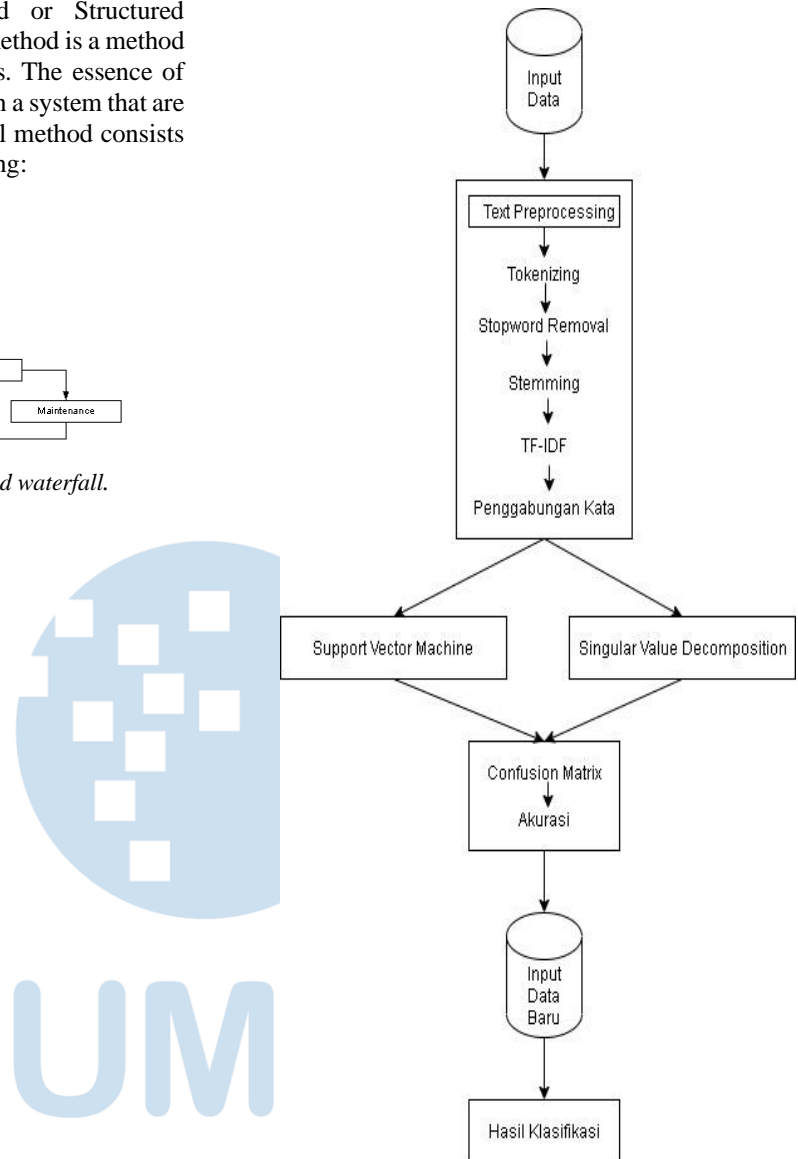


Figure 2. Diagram Block.

The classification process starts from input data in the form of news titles, then continues with text operations, in this process there are several stages, namely the tokenizing stage to separate words and convert them into spaces, stopword stages to delete words that do not contain meaning, stemming stages to remove affixed words and weighting or TF-IDF for the process of giving the index or frequency contained in the final word of the stemming process, then it will enter the word merging process (synonym), if there are different words but have the same meaning, then the system can combine them together with the frequency, Stages The next step is Support Vector Machine and Singular Value Decomposition. Next is the testing phase using the Confusion Matrix, the

number of correct predictions is divided by the total of all data. And the last one is Classification with new text data in the form of title and news content to determine the category of the news.

### C. Tokenizing

In the tokenizing the process that occurs is the process of breaking sentences into word for word, these words are also changed from uppercase to lowercase and eliminate unique characters that are not included in the word.

### D. Stopword Removal

The next process is Stopword, which is the process of filtered words where unimportant words from the text will be discarded. The system will check the Stopword list dictionary, if the word exists then the word will be deleted.

### E. Stemming

After the Stopword stage, the next process is the Stemming process. Where the system will search for words from the existing news text and convert them into basic words.

## IV. RESULT AND DISCUSSION

In this chapter, are the stages of implementation and testing of the system that has been built.

### A. User Interface

The image below is an image of the Process Data, display from the program that was created.

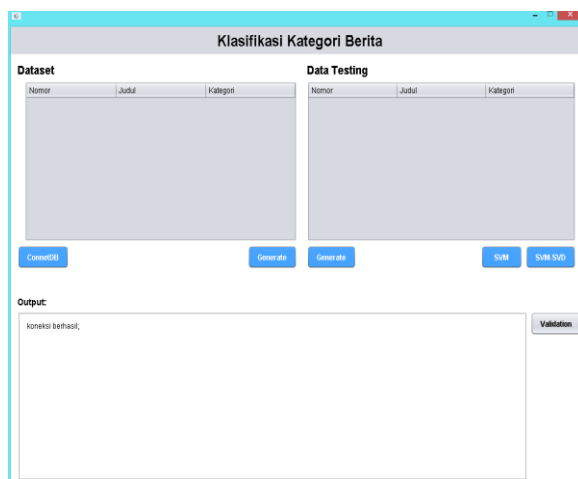


Figure 3. Process Data.

### News Classification

The image below is a display of the new news classification

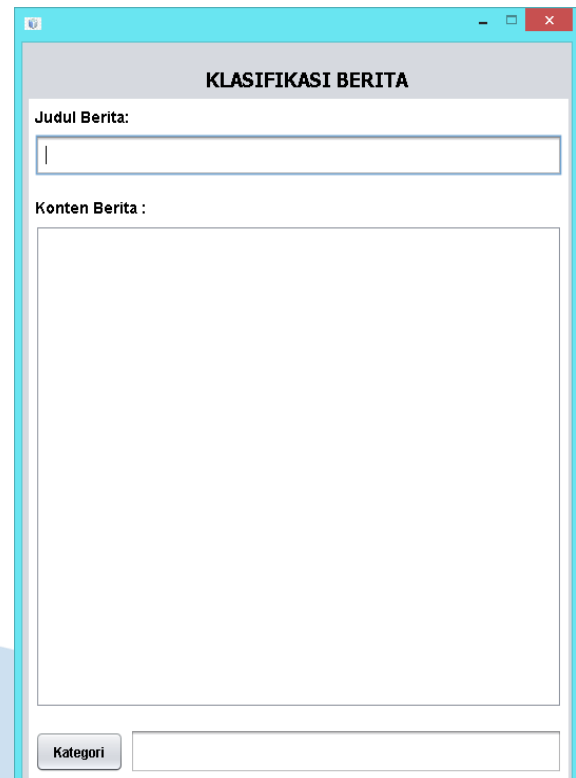


Figure 4. News Classification

### B. Support Vector Machine Algorithm Classification

In this stage, a news category classification process will be carried out based on the title and content of the news, there are 480 data on the title and content of the news with the division of 360 Training Data and 120 Testing Data, the following results are obtained:

Total Term TF-IDF 4092 from 480 data documents News In this process the system will carry out a machine learning and training process which only uses 1 algorithm, namely Support Vector Machine and accuracy calculations on the Confusion Matrix.

#### Train Data

```
dataBerita.addAll(train1);
dataBerita.addAll(train2);
kelas[0]=train1.get(0).getKategori();
kelas[1]=train2.get(0).getKategori();
data=new
double[dataBerita.size()][dataBerita.get(0).getSvd(
).length];
y=new int[dataBerita.size()];
alpha=new double[dataBerita.size()];
Arrays.fill(alpha, 0);
deltaAlpha=new double[dataBerita.size()];
Arrays.fill(deltaAlpha, 0);
error=new double[dataBerita.size()];
w=new double[dataBerita.size()];
int i;
for(i=0; i<dataBerita.size(); i++){
data[i]=dataBerita.get(i).getSvd();
if(dataBerita.get(i).getKategori().equals(kelas[0]))
```

```

y[i]=-1;
else y[i]=1; }
setKernel();

Set Kernel
kernel=new double[data.length][data.length];
int i,j;
for(i=0; i<data.length; i++){
for(j=0; j<data.length; j++){
kernel[i][j]=hitungKernel(data[i], data[j]); }}

Hitung Kernel
int row=x1.length;
int a=0;
double hasil=0;
while(a<row){
hasil+=x1[a]*x2[a];
++;}
hasil+=c;
hasil=Math.pow(hasil, d);
return hasil;
    
```

Figure 5. Source Code SVM

- The average value of the Accuracy Value is 81%
- The Average Value of Precision is 82%
- The average value of the recall value is 61%

Calculation f1-score

$$f1 = 2 \times \frac{82 \times 61}{82 + 61} = 69,96$$

From the calculation of the f1-score against the Support Vector Machine algorithm, a value of 69.96% is obtained, and with the Algorithm processing time for 35.81 seconds.

C. Classification of Support Vector Machine Algorithm with Singular Value Decomposition

Singular Value Decomposition is a technique used to decompose matrices of any size. to facilitate data processing. Here it is used with a K-Rank value of 50% then the results are as follows:

**Output:**

Confusion Matrix				
Kategori	Entertainment	Kesehatan	Politik	Teknologi
Entertainment	14	0	11	2
Kesehatan	0	6	23	2
Politik	0	0	31	0
Teknologi	0	0	8	23
TP:	14	6	31	23
FP:	0	0	42	4
FN:	13	25	0	8
TN:	93	89	47	85
Akurasi:	0,89	0,79	0,65	0,90
Presisi:	1,00	1,00	0,42	0,85
Recall:	0,52	0,19	1,00	0,74

Rata-rata	
Akurasi:	0,81
Presisi:	0,82
Recall:	0,61

Waktu Proses: 35,81 detik

Figure 6. Confusion Matrix SVM Algorithm.

From Figure 5 we get:

- The Accuracy Value for the Entertainment Category is 89%, the Health Category is 79%, the Political Category is 65% and the Technology Category is 90%.
- The Precision Value for the Entertainment Category is 100%, the Health Category is 100%, the Political Category is 42% and the Technology Category is 85%.
- Recall value for Entertainment Category is 52%, Health Category is 19%, Political Category is 100% and Technology Category is 74%.

And Average Score:

```

setRank
System.out.println(k);
if(k<1 || k>=Math.min(m, n)) return;
int i, j;
double[][] hasil;
hasil=new double[m][k];
for(i=0; i<m; i++){
for(j=0; j<k; j++){
hasil[i][j]=u[i][j]; }
u=hasil;

GetVector
double[] hasil=new double[s.length];
double[] temp=new double[s.length];
int i, j;
for(i=0; i<u[0].length; i++){
for(j=0; j<tfidf.length; j++){
temp[i]+=tfidf[j]*u[j][i]; } }
for(i=0; i<si.length; i++){
for(j=0; j<temp.length; j++){
hasil[i]+=temp[j]*si[j][i]; } }
return hasil;
u=hasil;
hasil=new double[k][k];
for(i=0; i<k; i++){
for(j=0; j<k; j++){
hasil[i][j]=s[i][j]; } }
s=hasil;
    
```

```

Matrix msi=new Matrix(s);
si=msi.inverse().getArrayCopy();
hasil=new double[n][k];
for(i=0; i<n; i++){
for(j=0; j<k; j++){
hasil[i][j]=v[i][j]; }}
v=hasil;

```

Figure 7. Source Code SVD.

**Jumlah term TF-IDF:** 4092  
**Hasil reduksi fitur SVD:** 180  
**Penggunaan K-rank SVD:** 50%

Figure 8. Results of Using K-Rank SVD.

**Output:**

Confusion Matrix

Kategori	Entertainment	Kesehatan	Politik	Teknologi
Entertainment	26	0	1	0
Kesehatan	5	23	3	0
Politik	1	0	30	0
Teknologi	1	0	3	27

TP: 26      23      30      27  
FP: 7      0      7      0  
FN: 1      8      1      4  
TN: 86      89      82      89

Akurasi: 0,93      0,93      0,93      0,97  
Presisi: 0,79      1,00      0,81      1,00  
Recall: 0,96      0,74      0,97      0,87

Rata-rata  
Akurasi: 0,94  
Presisi: 0,90  
Recall: 0,89

Waktu Proses: 29,22 detik

Figure 9. Confusion Matrix Results of SVM and SVD Algorithms.

From Figure 7 we get:

- Accuracy Value for Entertainment Category is 93%, Health Category is 97%, Political Category is 93% and Technology Category is 97%.
- The Precision Value for the Entertainment Category is 79%, the Health Category is 100%, the Political Category is 81% and the Technology Category is 100%.
- Recall value for Entertainment Category is 96%, Health Category is 74%, Political Category is 97% and Technology Category is 87%.

And Average Score:

- The average value of the Accuracy Value is 94%
- Average Precision Value of 90%

- The average value of the recall value is 89%
- Calculation *f1-score*

$$f1 = 2 \times \frac{90 \times 89}{90 + 89} = 89,50$$

From the calculation of *f1-score* against the Support Vector Machine algorithm with Singular Value Decomposition, the value is 89.50%, and with the Algorithm processing time for 29.22 seconds.

#### D. Comparison of SVM and SVM-SVD Hasil Results

Table 2. Comparison of SVM and SVM-SVD Results

	Support Vector Machine				
	E	K	P	T	Average
Accuracy	89	79	65	90	<b>81</b>
Precision	100	100	42	85	<b>82</b>
Recall	52	18	100	74	<b>61</b>
Support Vector Machine dan Singular Value Decomposition					
	E	K	P	T	Average
Accuracy	93	93	92	97	<b>94</b>
Precision	79	100	81	100	<b>90</b>
Recall	96	74	97	87	<b>89</b>

From table 2, the comparison between the Support Vector Machine Algorithm and the Singular Value Decomposition results in a higher level of Precision, Recall and Accuracy using the Support Vector Machine Algorithm with the addition of the Singular Value Decomposition Method rather than using only one Support Vector Machine Algorithm. And also for the time required for the classification process on the Support Vector Machine Algorithm with the Singular Value decomposition method, the classification duration is faster with a time of 29.22 seconds compared to only using the Support Vector Machine Algorithm with a time of 35.81 seconds.

#### E. New Data Classification

At this stage, data validation will be carried out by entering New Data in the form of News Title and Contents then the system will automatically detect the News Category

Experiment with new data.

**KLASIFIKASI BERITA**

**Judul Berita:**  
Red Velvet Cake Disebut Jadi Nama Baru Android 11

**Konten Berita :**

Fitur-fitur Android 11 versi beta telah dirilis oleh The Verge dan cukup akrab bagi pengguna iOS, di antaranya:

1. OS-Level SmartHome Control

Salah satu perubahan yang paling mencolok di Android 11 ialah tersedianya fitur kontrol rumah pintar yang dikembangkan di OS Android.

Sebelumnya, Apple telah lebih dulu menambahkan fitur HomeKit ke menu Control Center Apple di iOS 10.

Bagi Anda pengguna ponsel Android, fitur ini memudahkan Anda untuk mengontrol perangkat pintar yang terhubung ke Google Home di mana saja

**Kategori** Teknologi

Figure 10. New Data Classification.

The results of experiments with News Title and Content are classified as Technology category.

From the results of the classification in Figure 8, it is known the accuracy of each category. The accuracy of each category is obtained from the number of words/terms after passing through the Text Preprocessing stages. The existing words/terms will be checked into the Training Data and Testing Data which will then be calculated for the total Words/Terms contained in each news category in the existing Training Data and Testing Data.

Tingkat Persentasi Masing-Masing Kategori

Entertainment	6,90 %	2 Kata
Kesehatan	6,90 %	2 Kata
Teknologi	75,86 %	22 Kata
Politik	10,34 %	3 Kata
Total	100 %	29 Kata

Berita Tersebut Masuk Kategori Teknologi dengan Akurasi 75,86 %

Figure 11. Accuracy of each Category.

## V. CONCLUSION

The conclusions that can be drawn from the development of the News classification recommendation system with the Support Vector Machine Algorithm and Singular Value Decomposition are as follows:

- 1) The accuracy level of the Support Vector Machine Algorithm from Experiments with a total of 360 Training Data and 120 Testing Data obtained an Accuracy Value of 81%
- 2) The Accuracy Level of the Support Vector Machine Algorithm and Singular Value Decomposition from Experiments with a total of 360 Training Data and 120 Testing Data obtained an Accuracy Value of 94%
- 3) From the results of experiments conducted Singular Value Decomposition can be applied with a Support Vector Machine and there is an increase in accuracy results and faster processing time.
- 4) From the results of the f1-score calculation, the Support Vector Machine Algorithm and Singular Value Decomposition values are higher by 89.50% compared to using the Support Vector Machine Algorithm only with a value of 69.96%.

The Support Vector Machine and Singular Value Decomposition Algorithms in this study were only tested with a little data, it would be better if there were more data.

## REFERENCES

- [1] I. S. Widodo, S. Pd, M. Si, P. Kasih, and M. Kom, "Klasifikasi Konten berita Surat Kabar Berdasarkan Judul Dengan Text Mining Menggunakan Metode Naive Bayes (STUDI KASUS : RADAR KEDIRI ) Oleh : Enggal Suci Febriani Dibimbing oleh : Surat Pernyataan Artikel Skripsi Tahun 2017," 2017.
- [2] B. Kurniawan, S. Effendi, and O. S. Sitompul, "Klasifikasi Konten Berita Dengan Metode Text Mining," *J. Dunia Teknol. Inf.*, vol. 1, no. 1, pp. 14–19, 2012.
- [3] D. Ariadi and K. Fithriasari, "Klasifikasi Berita Indonesia Menggunakan Metode Naive Bayesian Classification dan Support Vector Machine dengan Confix Stripping Stemmer," *J. SAINS DAN SENI ITS Vol. 4, No.2*, vol. 4, no. 2, pp. 248–253, 2015.
- [4] A. S. Sri Widaningsi, "Klasifikasi Jurnal Ilmu Komputer Berdasarkan Pembagian," *Semin. Nas. Teknol. Inf. dan Komun. 2018 (SENTIKA 2018)*, vol. 2018, no. Sentika, pp. 23–24, 2018.
- [5] G. Ngurah, M. Nata, and P. P. Yudiastra, "Preprocessing Text Mining Pada Email Box Berbahasa Indonesia," *Konf. Nas. Sist. Inform. 2017*, pp. 479–483, 2017.
- [6] V. Amrizal, "Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim)," *J. Tek. Inform.*, vol. 11, no. 2, pp. 149–164, 2018, doi: 10.15408/jti.v11i2.8623.
- [7] B. Herwijayanti, D. E. Ratnawati, and L. Muflikhah, "Klasifikasi Berita Online dengan menggunakan Pembobotan TF-IDF dan Cosine Similarity," *Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 1, pp. 306–312, 2018.
- [8] A. H. Setianingrum, D. H. Kalokasari, and I. M. Shofi, "Implementasi Algoritma Multinomial Naive Bayes Classifier," *J. Tek. Inform.*, vol. 10, no. 2, pp. 109–118, 2018, doi: 10.15408/jti.v10i2.6822.
- [9] M. A. Fatmawati, "Klasifikasi Keluhan," *J. CoreIT*, vol. 3, no. 1, pp. 24–30, 2017, doi: 10.1007/s00766-002-0156-7.
- [10] A. Setiyono and H. F. Pardede, "Klasifikasi SMS Spam Menggunakan Support Vector Machine," *J. PILAR Nusa*



- Mandiri*, vol. 15, no. 2, pp. 275–280, 2019, doi: 10.33480/pilar.v15i2.693.
- [11] U. I. N. Ar-raniry and B. Aceh, “Analisis dan perbandingan kualitas pengelompokan dokumen ( document clustering ) dengan menggunakan metode k-means dan k-medians bustami yusuf,” vol. 1, no. 2, pp. 1–10, 2015.
- [12] V. Amrizal, “Penerapan Metode Term Frequency Inverse Document Frequency (Tf-Idf) Dan Cosine Similarity Pada Sistem Temu Kembali Informasi Untuk Mengetahui Syarah Hadits Berbasis Web (Studi Kasus: Hadits Shahih Bukhari-Muslim),” *J. Tek. Inform.*, vol. 11, no. 2, pp. 149–164, 2018, doi: 10.15408/jti.v11i2.8623.

