

## Identification of Hoax News in the Using Community TF-RF and C5.0 Tree Decision Algorithm

Enrico Budi Santoso<sup>1\*</sup>, Yulison Herry Chrisnanto<sup>2</sup>, Gunawan Abdillah<sup>3</sup>

<sup>1\*,2,3</sup> Universitas Jenderal Achmad Yani, Indonesia

\*Email: rico.budi25@gmail.com, y.chrisnanto@gmail.com

ARTICLE INFO	ABSTRACT
<p><b>Keywords:</b> News Hoax Decision Tree C5.0 TF.RF Data Mining</p>	<p><i>News has a great influence on social and political conditions, and news can drive the economy of a country. Identifying hoax news is very important to ensure that the information circulating in society is true and reliable, and helps limit the spread of false information. In the process of reading news spread on social media, people do not know whether it is fact or hoax news because they cannot distinguish whether the news circulating is real news or fake news which if left unchecked can result in the public being misinformed. Therefore, this research process is to create a sistem for identifying hoax news using Decision Tree C5.0, which is an algorithm for the development of the C4.5 algorithm which in a process is almost similar, but the C5.0 algorithm has more value than the C4.5 algorithm which is used for the data mining process with a classification method for 1000 data obtained by web scraping using the keywords "election 2024", "politics" and "checkfaktapilkadamafindo" on the Turnbackhoax.id and Detik.com sites. In this study, what distinguishes it from several previous studies is its existence in several test scenarios, namely classification using feature weighting, which in classification using feature weighting is TF.RF. After testing the confusion matrix on the C5.0 algorithm, it produces accuracy, precision, and recall on each training / test data (70/30) resulting in accuracy 79.33%, precision 80.00%, recall 97.00%, then training / test data (80/20) resulting in accuracy 79.50%, precision 81.00%, recall 95.00%, then training and test data (90/10) resulting in accuracy 72.00%, precision 74.00%, recall 89.00%.</i></p>

### INTRODUCTION

Hoax news has a strong impact on the reporting of information, especially on social media (Handika et al., 2022). In this era of rapid progress in social media, reporting of information is very fast if hoax reporting is left unchecked, which can lead to people being misinformed by news received (Utomo et al., 2020). For example, fake news about politics can really disrupt social life in society if it is not prevented (Ghaniaviyanto Ramadhan et al., 2022). In an effort to anticipate the spread of hoax news, this is by classifying the news as to whether the news circulating is genuine news. or is fake news so that it can provide an explanation to readers of the news (Granik & Mesyura, 2017). The detection of hoax news consistently increases from year to year, which means that efforts to stem the circulation of hoax news never end and continue to be implemented (Prayoga et al., 2021a).

According to information released by the Ministry of Communication and Information from 2018 to 2019, there were 3356 hoax news that spread on social media, and the most about politics, namely 916 hoax news (Beta et al., 2022). Hoax news can cause people to frequently difficulty in sorting out what is valid and what is a hoax on social media. If you are not alert, people will easily be influenced by hoax news and will participate in spreading the hoax news. Therefore, an algorithm is needed to classify news whether it is valid or a hoax (Wati, 2020).

Scientists explain that classifying a dataset is using a machine learning method which is an effective method for classifying data in the form of hoax news (Prayoga et al., 2021b). One of the machine learning algorithms is the C5.0 algorithm which is a development algorithm from the C4.5 algorithm where the process is almost similar, but

the C5.0 algorithm has more value than the C4.5 algorithm. The result of the C5.0 algorithm is a decision tree that is created based on the entropy or gain value which can be used for classification process problems (Utami et al., 2020). Then in the word weighting process using the TF.RF method, which is a combination of the Term Frequency method and the Relevance Frequency method. Combining these two methods can produce good performance (Sains et al., 2022a).

In this research, we will apply a classification method using the C5.0 algorithm and TR.RF word weighting in the process of identifying hoax news using accuracy calculations referring to the confusion matrix. The system that was built also aims to find out what features have an influence on identifying hoax news and find out the performance level of the C5.0 algorithm in the process of classifying news that is considered a hoax in society.

## **METHOD**

In this research there are several processes in achieving results that are in line with what is expected, the first process is collecting news datasets, the second process is data preprocessing which consists of four stages, then the third process is word weighting, then the fourth process is dividing training/test data, then the fifth process carries out classification using the C5.0 decision tree, and finally the sixth process is carrying out evaluation using the confusion matrix.

### **News Dataset**

The news dataset was collected from web scraping using the keywords "2024 election", "politics" and "cekkfakpilkadamafindo" on the Turnbackhoax and Detik websites from 2023 to 2018. The data collection results obtained 1000 political news stories which were then stored in the CSV extension. The information obtained from web scraping is the title, article content, date, and link. Fact data taken from the website <https://www.detik.com>. News titles are obtained from the text in the <h1> tag, namely 'title' and news articles are obtained in the <p> tag, namely 'narrative'. The data obtained is then placed using the name DataFakta using the CSV extension. Hoax data is taken on the website <https://turnbackhoax.id>. The news title is taken from the text in the <h1> tag, namely 'title' and the news article is obtained in the <p> tag, namely 'narrative'. The data obtained is then placed using the name DataHoax with the CSV extension. The data that has been collected is then combined into news data where the data will be entered into the system in CSV format and then stored in the database.

### **Preprocessing**

The text preprocessing stage is a stage in preparing raw data before proceeding to the next stage. At this stage, we clean the data to combine word forms and reduce the number of words in the document collection.

### **TF-RF Weighting**

The weighting method uses TF-RF, a development of TF-IDF, which is refined by using relevance frequency (RF), which is expected to increase discriminatory power (Willy et al., 2019). TF-RF is a combined method between TF and RF that considers the relevance of documents as seen from the frequency of occurrence of words in that category and aims to obtain a better level of performance compared to other word weighting methods (Deolika et al., 2019).

### **Dataset Division**

In this process, the data that has been combined will be divided into training data and then testing data which is carried out by the system. The value of the data sharing ratio is tested through several scenarios to understand the ratio of training data to testing data with several models so that they can perform well. In this research, there is a data separation scenario, namely: training/testing data (70:30), training/testing data (80:20), and training/testing data (90:10) (Irena & Erwin Budi Setiawan, 2020).

### **Classification Using the C5.0 Algorithm**

The research will be carried out using the C5.0 algorithm which has a resulting tree with a varying number of branches at each node. C5.0 treats a continuous variable the same as CART, but in categorical variables, C5.0 treats a categorical variable value as a splitter (Agustina et al., 2021). The C5.0 algorithm is an update of C4.5 where C5.0 is better than the C4.5 algorithm in terms of time, memory usage efficiency, decision tree scale, and deviation in classification. In the C5.0 algorithm, the initial training data is recognized as the root node in the decision tree, then calculates the gain ratio for each attribute (Meo, 2017).

## Learning Model

The learning model is a system that has been created from the C5.0 decision tree classification. The testing data for which the scenario has been created will be tested on the model that has been built. The output created produces True/False, at which stage a model can produce the results of the classification process sourced from testing data, where the results of the classification are documents that are classified as false, namely label 1, and true classification, namely label 0.

## Evaluation Stage

In this research, the model evaluation process uses the Confusion Matrix technique in the process of obtaining model accuracy in Data Mining. This accuracy is used in the accuracy identification process

## RESULTS AND DISCUSSION

### Implementation

In this implementation, it explains the software implementation process that is implemented according to the software design explained previously. What is discussed in the implementation is software implementation using the Javascript programming language which is run using the Visual Studio application, while implementing the database using MongoDB, then to display the program using a web browser such as Google Chrome.

#### Database Implementation

The database implementation process is carried out based on the database design analysis process that has been carried out previously. The database for this system is done using MongoDB. The database is made of three tables and has several attributes that support the running of the system created, namely the Hoax News Identification System in Society Using TF-RF and the C5.0 Decision Tree Algorithm. To further explain the database implementation process, see below.

#### 1. News Table

The news table is implemented based on the database design analysis that has been previously designed. The implementation of the new table can be seen in Figure 1.

#	_id ObjectId	id String	label String	tanggal String	judul Stri
1	ObjectId('64898ed5e3a7fa1d985...')	"71"	"1"	"22-Jul-23"	"PDIP usung"
2	ObjectId('64898ed5e3a7fa1d985...')	"461"	"0"	"24-Jan-23"	"Disindir D"
3	ObjectId('64898ed5e3a7fa1d985...')	"495"	"0"	"24-Jan-23"	"Pakar Nila"
4	ObjectId('64898ed5e3a7fa1d985...')	"550"	"0"	"23-Jan-23"	"Prabowo Ya"
5	ObjectId('64898ed5e3a7fa1d985...')	"681"	"0"	"23-Jan-23"	"Prabowo da"
6	ObjectId('64898ed5e3a7fa1d985...')	"736"	"0"	"23-Jan-23"	"Dukungan J"
7	ObjectId('64898ed5e3a7fa1d985...')	"1161"	"0"	"23-Jan-23"	"Survei Alg"
8	ObjectId('64898ed5e3a7fa1d985...')	"1597"	"0"	"21-Jan-23"	"Fahri Hamz"
9	ObjectId('64898ed5e3a7fa1d985...')	"2098"	"0"	"21-Jan-23"	"Bos Parame"
10	ObjectId('64898ed5e3a7fa1d985...')	"2226"	"1"	"20-Jan-23"	"JOKOWI STR"
11	ObjectId('64898ed5e3a7fa1d985...')	"2588"	"1"	"20-Jan-23"	"KPK GELEDA"

Figure 1. Implementation of the News Table Database

#### 2. Case Folding Table

The case folding table is implemented based on an analysis of the database design that has been previously designed. The implementation of the case folding table can be seen in Figure 2.

casefoldings							
_id	ObjectId	idberita	String	narasi	String	_v	Int32
1	ObjectId('6474012a8f72322d983...	"6468d6153651bf261cb5fa58"	"dalam isi video tersebut tid...	0			
2	ObjectId('6474012a8f72322d983...	"6468d6153651bf261cb5fa59"	"partai demokrat (pd) menyind...	0			
3	ObjectId('6474012a8f72322d983...	"6468d6153651bf261cb5fa5a"	"partai gerindra dan partai k...	0			
4	ObjectId('6474012a8f72322d983...	"6468d6153651bf261cb5fa5b"	"ketua umum gerindra prabowo ...	0			
5	ObjectId('6474012a8f72322d983...	"6468d6153651bf261cb5fa5c"	"ketua umum (ketum) partai ge...	0			
6	ObjectId('6474012a8f72322d983...	"6468d6153651bf261cb5fa5d"	"nama ganjar pranowo, parbowo...	0			
7	ObjectId('6474012a8f72322d983...	"6468d6153651bf261cb5fa5e"	"algoritma research & consult...	0			
8	ObjectId('6474012a8f72322d983...	"6468d6153651bf261cb5fa5f"	"saling adu pernyataan politi...	0			
9	ObjectId('6474012a8f72322d983...	"6468d6153651bf261cb5fa60"	"wali kota solo gibran rakabu...	0			
10	ObjectId('6474012a8f72322d983...	"6468d6153651bf261cb5fa61"	"klain gibran rakabuming raka...	0			
11	ObjectId('6474012a8f72322d983...	"6468d6153651bf261cb5fa62"	"pengeledahan yang dilakuka...	0			
12	ObjectId('6474012a8f72322d983...	"6468d6153651bf261cb5fa63"	"selain berisi gambar pratinj...	0			

Figure 2 Database Implementation of Case Folding Tables

### 3. Results Table

The results table is implemented based on the database design analysis that has been previously designed. The implementation of the results table can be seen in Figure 3.

hasil									
_id	ObjectId	id	String	judul	String	narasi	String	label	String
1	ObjectId('649bd508a547678ac5...	""	"Disindir Demokrat, Ini Alasa...	"Partai Demokrat (PD) menyind...	"0"				
2	ObjectId('649bd508a547678ac5...	""	"PDIP usung Megawati di pilpr...	"Dalam isi video tersebut tid...	"1"				
3	ObjectId('649d99030c0ae929def...	""	"Demokrat Kejang2 SBY AHY Gak...	"Faktanya tidak ada pemberita...	"1"				
4	ObjectId('649d99c08c0ae929def...	""	"Pakar Nilai Peresmian Sekber...	"Partai Gerindra dan Partai K...	"0"				
5	ObjectId('649d9b08c0ae929def...	""	"Ganjar calon presiden 2024"	"Ganjar calon presiden 2024"	"0"				
6	ObjectId('64a03d17caadf2d8ef...	""	"Demokrat Kejang2 SBY AHY Gak...	"Faktanya tidak ada pemberita...	"1"				

Figure 3 Database Implementation Results Table

### Interface Implementation

In the process of implementing the interface for the Hoax News Identification System in Society Using TF-RF and the C5.0 Decision Tree Algorithm, it was created based on the system interface design analysis process that had been carried out previously for the program created using JavaScript which was run using Visual Studio (Hendriyanto & Sari, 2022a). The following are the interface implementation stages, At this stage, the system is in display form and can be operated or executed according to the design results.

#### 1. Implementation of the Dashboard Page

The following displays the dashboard page which will be accessed first by the user when the user enters the Hoax News Identification System in Society Using TF-RF and the C5.0 Decision Tree Algorithm, for The implementation of the dashboard page can be seen in Figure 4.

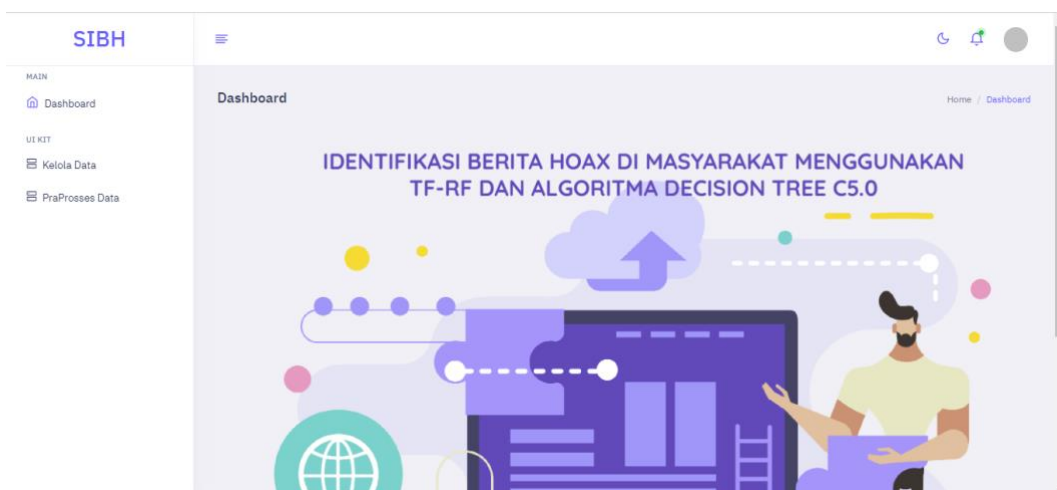


Figure 4 Dashboard Page Implementation

## 2. Implementation of the Manage Data Page

The following is the display for the data management page which the user will access when importing and viewing previously imported datasets where the dataset will be managed by systems such as preprocessing (Hendriyanto & Sari, 2022b), weighting, classification, and accuracy testing, for implementation of the data management page can be seen at Figure 5.

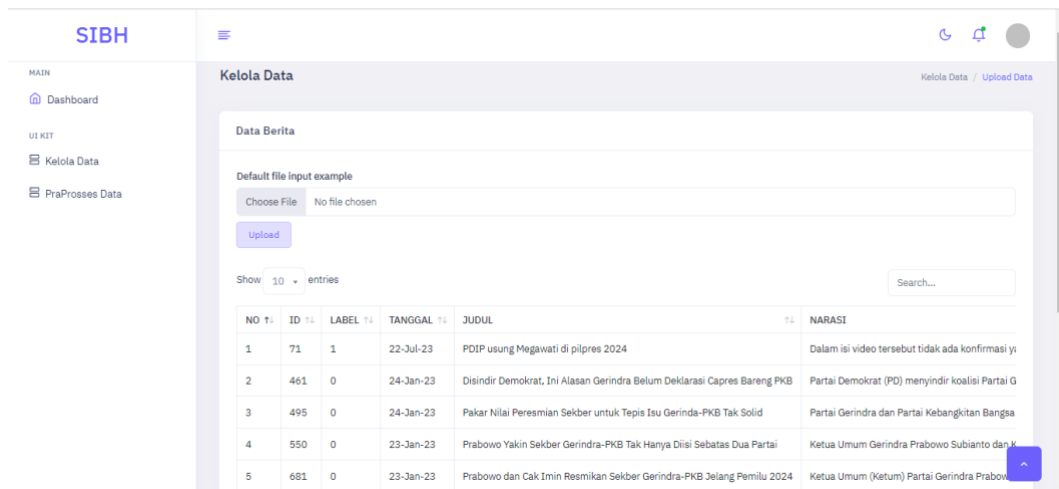


Figure 5 Implementation of the Manage Data Page

## 3. Preprocess Page Implementation

The following is the display for the preprocessing page which will be accessed by the user when the user before carrying out the case folding process this page first (Sains et al., 2022b), for the implementation of the data preprocessing page, can be seen in Figure 6.

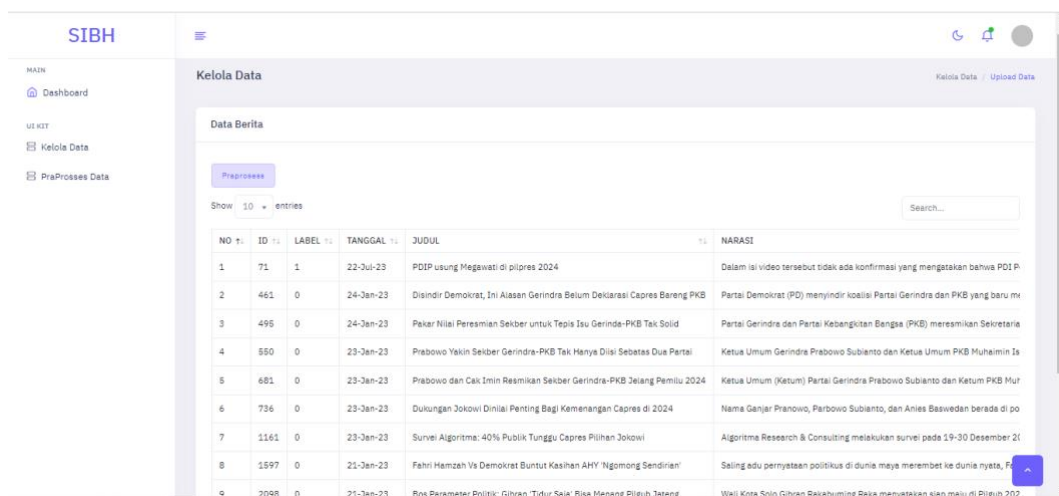


Figure 6. Preprocessing Page Implementation

## 4. Implementation of Case Folding Pages

The following is the display for the case folding page which will be accessed by the user when the user has carried out the case folding process. The results of the case folding will be displayed in a table, for the implementation of the case folding page can be seen in Figure 7.

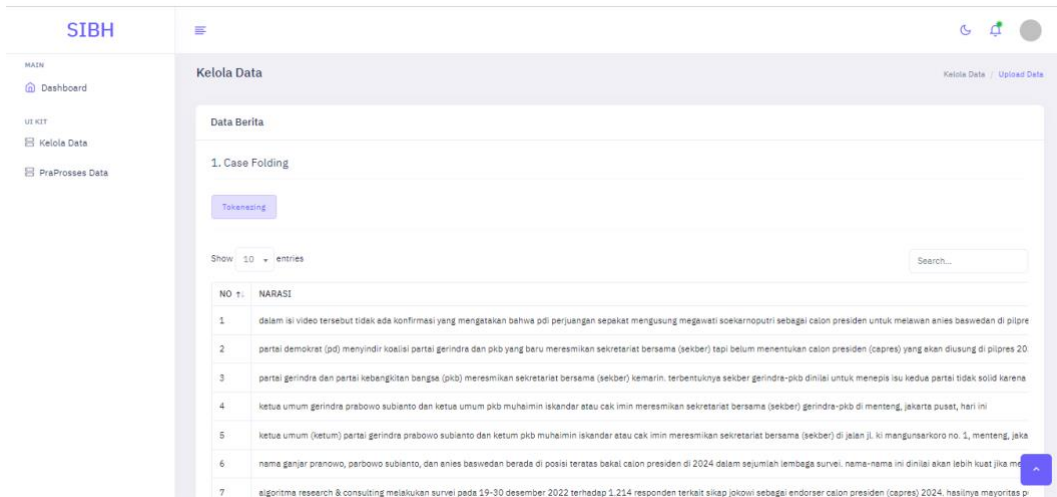


Figure 7 Implementation of Case Folding Pages

### 5. Implementation of Tokenizing Pages

The following displays the tokenizing page which will be accessed by the user when the user has carried out the tokenizing process, The results of the tokenizing will be displayed in a table, for the implementation of the tokenizing page can be seen in Figure 8.

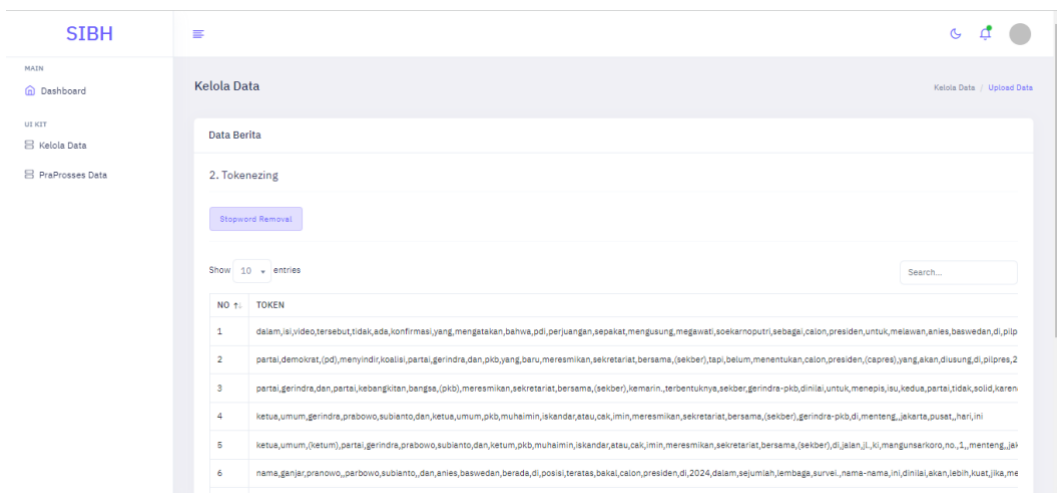


Figure 8 Implementation of Tokenizing Page

### 6. Implementation of Stopword Removal Page

The following is the display for the stopword removal page which will be accessed by the user when the user has carried out the stopword removal process, The results of the stopword removal will be displayed in a table, for the implementation of the stopword removal page can be seen in Figure 9.

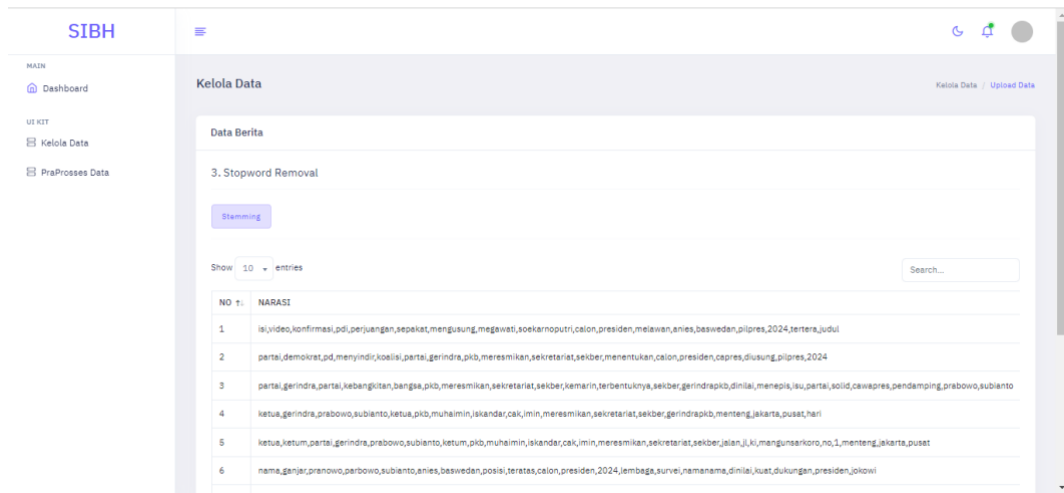


Figure 9 Implementation of the Stopword Removal Page

### 7. Stemming Page Implementation

The following is the display for the stemming page which will be accessed by the user when the user has carried out the stemming process. The results of the stemming will be displayed in a table, for the implementation of the stemming page can be seen in Figure 10.

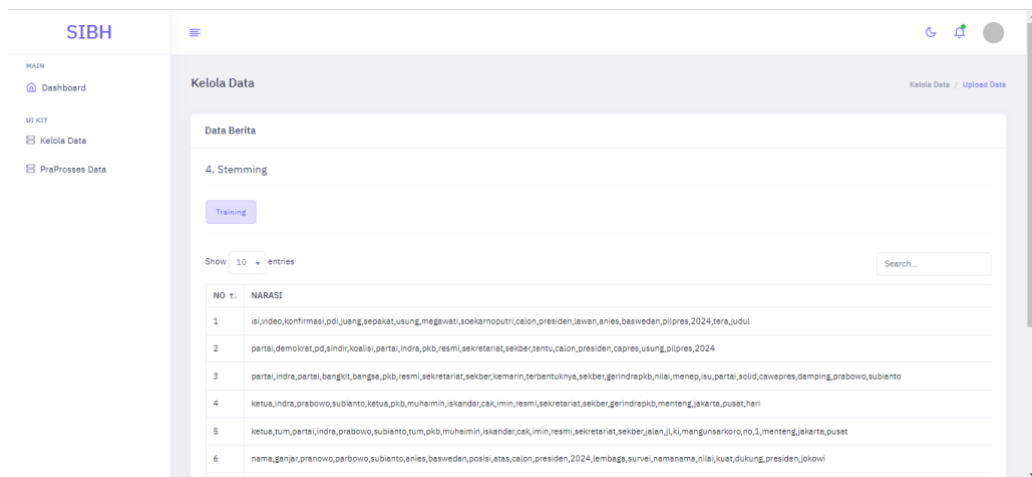


Figure 10 Stemming Page Implementation

### 8. Train Page Implementation

The following is the display for the training page which consists of a validation test, news input, and a view of the decision tree which will be accessed by the user after completing training which produces an accuracy test while at the bottom there is a news input form which is used for news identification, then there is also a tree decision which are the results of classification, for the implementation of validation test pages, news input, see the decision trees in Figure 11, Figure 12, and Figure 13.

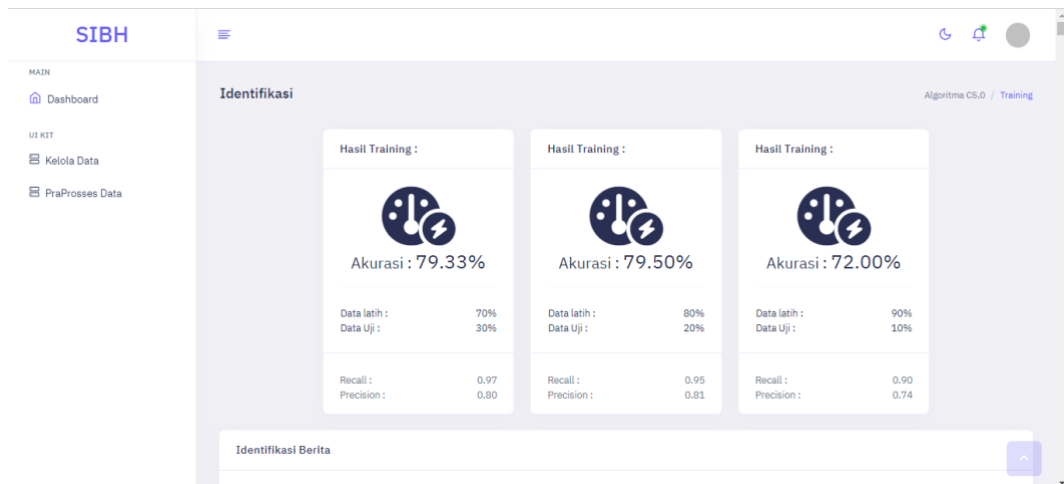


Figure 11 Implementation of the Accuracy Test Page

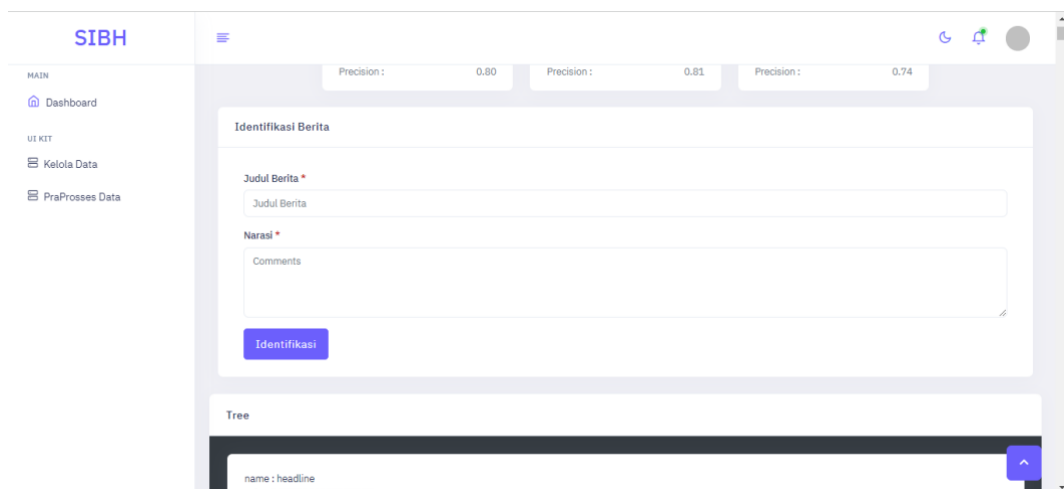


Figure 12 Implementation of the News Input Page

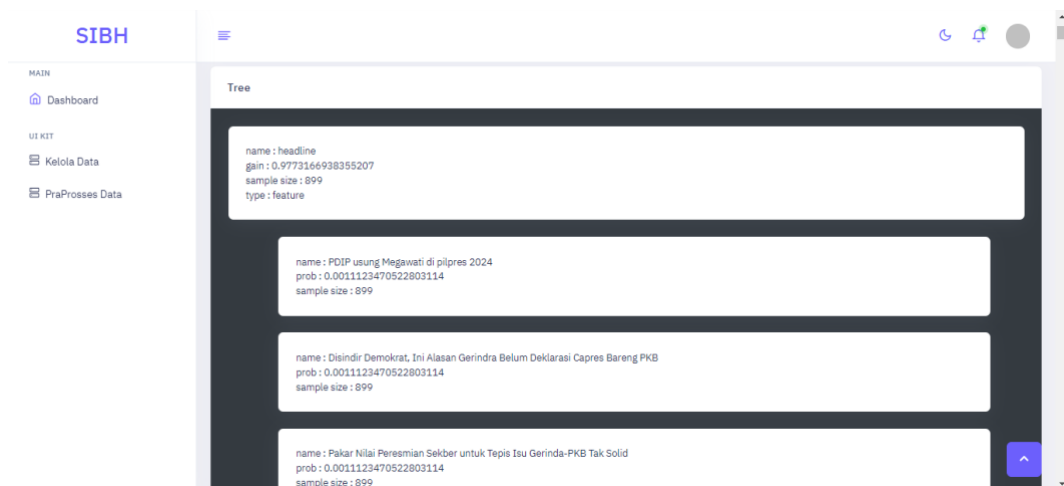


Figure 13 Implementation of Decision Tree View

## 9. Implementation of the Classification Results Page

The following is the display for the classification results page which will be accessed by the user after completing the input of news which will be classified to identify the news whether the title and narrative are hoaxes or valid. Then the results are displayed in a table, for the implementation of the classification results page can be seen in Figure 14.

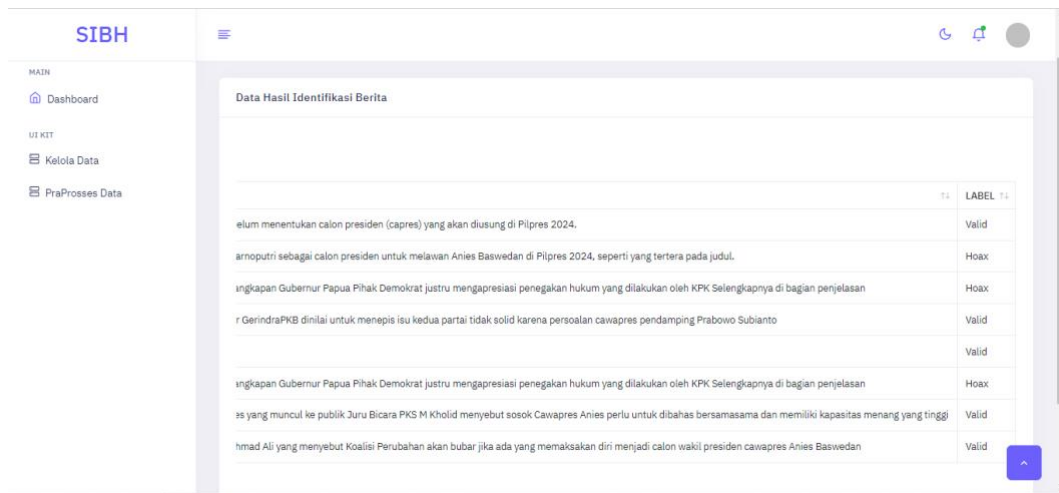


Figure 14 Implementation of the Classification Results Page

### Algorithm Implementation

The following is an implementation of the C5.0 algorithm for the hoax news classification process by selecting the attribute that has the highest gain value and then choosing it to be the parent of the next node. The following is an implementation of the algorithm as in Figure 15.

```

421 // Algoritma C5.0
422 // Fungsi untuk menghitung entropi
423 function entropy(data) {
424   const labelCounts = {};
425   for (let i = 0; i < data.length; i++) {
426     const label = data[i].label;
427     if (!labelCounts[label]) {
428       labelCounts[label] = 0;
429     }
430     labelCounts[label]++;
431   }
432
433   let ent = 0;
434   const dataLength = data.length;
435   for (const label in labelCounts) {
436     const p = labelCounts[label] / dataLength;
437     ent += p * Math.log2(p);
438   }
439   return ent;
440 }
441
442
443 // Fungsi untuk membagi data berdasarkan fitur dan nilai
444 function partitionData(data, feature, value) {
445   const trueSet = [];
446   const falseSet = [];
447   for (let i = 0; i < data.length; i++) {
448     const example = data[i];
449     if (example[feature] === value) {
450       trueSet.push(example);
451     } else {
452       falseSet.push(example);

```

Figure 15 Algorithm Implementation

### Confusion Matrix Test Results

In this research, the model evaluation process uses the Confusion Matrix technique in the process of obtaining model accuracy in Data Mining. This accuracy is used in the accuracy identification process. The indicators that will be calculated in the process of obtaining accuracy include:

- 1) True Positive (TP) is the total number of positive cases included in the correct classification.
- 2) True Negative (TN) total negative cases including correct classification.
- 3) False Positive (FP) is the total number of positive cases included in the wrong classification.
- 4) False Negative (FN) is the total number of negative cases included in the wrong classification.

### Calculating Confusion Matrix

The following is a way to calculate the confusion matrix based on the new data available in the image:

1. Calculating Confusion Matrix Training/Test Data (70/30)

```

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL

Confusion Matrix:
          | Predicted 1 | Predicted 0
Actual 1  | 227         | 7
Actual 0  | 55          | 11

Akurasi: 0.7933333333333333
Recall: 0.9700854700854701
Presisi: 0.8049645390070922
[ [ 227, 11, 55, 7 ] ]
    
```

Figure 16 Confusion Matrix (70/30)

- a. Calculating Accuracy =  $(TP + TN) / (TP + TN + FP + FN)$   
 Accuracy =  $(227+11)/(227+11+55+7)$   
 =  $0.7933 \times 100\%$   
 = 79.33%
- b. Calculating Precision = Precision:  $TP / (TP + FP)$   
 Precision =  $227/(227+55)$   
 =  $0.80 \times 100\%$   
 = 80.00%
- c. Calculating Recall = Recall =  $TP / (TP + FN)$   
 Recall =  $227/(227+7)$   
 =  $0.97 \times 100\%$   
 = 97.00%

2. Calculating Confusion Matrix Training/Test Data (80/20)

```

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL

Confusion Matrix:
          | Predicted 1 | Predicted 0
Actual 1  | 148         | 7
Actual 0  | 34          | 11

Akurasi: 0.795
Recall: 0.9548387096774194
Presisi: 0.8131868131868132
    
```

Figure 17 Confusion Matrix (80/20)

- a. Calculating Accuracy =  $(TP + TN) / (TP + TN + FP + FN)$   
 Accuracy =  $(148+11)/(148+11+7+34)$   
 =  $0.795 \times 100\%$   
 = 79.50%
- b. Calculating Precision = Precision:  $TP / (TP + FP)$   
 Precision =  $148/(148+7)$   
 =  $0.95 \times 100\%$   
 = 81.00%
- c. Calculating Recall = Recall =  $TP / (TP + FN)$   
 Recalls =  $148/(148+34)$   
 =  $0.81 \times 100\%$   
 = 95.00%

### 3. Calculating Confusion Matrix Training/Test Data (90/10)

```

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL

Confusion Matrix:
          | Predicted 1 | Predicted 0 |
Actual 1  | 61          | 7           |
Actual 0  | 21          | 11          |

Akurasi: 0.72
Recall: 0.8970588235294118
Presisi: 0.7439024390243902
[ [ 227, 11, 55, 7 ], [ 148, 11, 34, 7 ], [ 61, 11, 21, 7 ] ]
    
```

Figure 18 Confusion Matrix (90/10)

- a. Calculating Accuracy =  $(TP + TN) / (TP + TN + FP + FN)$   
 Accuracy =  $(61+11)/(61+11+7+21)$   
 =  $0.72 \times 100\%$   
 = 72.00%
- b. Calculating Precision = Precision:  $TP / (TP + FP)$   
 Precision =  $61/(61+7)$   
 =  $0.89 \times 100\%$   
 = 74.00%
- c. Calculating Recall = Recall =  $TP / (TP + FN)$   
 Recall =  $61/(61+21)$   
 =  $0.74 \times 100\%$   
 = 89.00%

Based on the results of accuracy testing carried out by the system using a confusion matrix where object 1 is valid while 0 is a hoax, then the conclusion is that the Hoax News Identification System in Society Using TF-RF and the C5.0 Decision Tree Algorithm produces accuracy, precision and recall on each training/test data (70/30) resulting in an accuracy of 79.33%, a precision of 80.00%, recall 97.00%, then training/test data (80/20) produces 79.50% accuracy, 81.00% precision, 95.00% recall, then training and test data (90/10), namely accuracy 72.00%, precision 74.00%, recall 89.00%.

#### Black Box Testing

Software testing aims to see the suitability between the design results and the expected goals. The purpose of quality testing is to determine the category of success in testing, designing quality testing, implementing testing, and concluding from the results of quality testing. The quality testing technique for the system in this research is carried out in the form of Black Box Testing which focuses on testing the functionality of the software.

#### Testing Stages

The testing stages carried out on the software in implementing this final assignment include the following activities:

1. Grouping processes based on the Use Cases that have been created.
2. Determine the objectives of quality testing.
3. Determine the category of quality test results.
4. Designing quality testing.
5. Implementation of quality testing.
6. Conclusion from the quality testing results.

**Testing Objectives**

This stage discusses the objectives of quality testing of the software being built, which can be seen in Table 1.

Table 1 Testing Objectives

No	Use Case	Purpose
1	Dataset	Carry out the dataset input process and see the dataset where the data will later be used for the preprocessing stage
3	Preprocessing	Preprocessing training data such as case folding, tokenizing, stopword removal, and stemming where the data can be used for further processing
4	Data Classification	Carry out the process of classifying data in the form of news whether it is hoax or valid news using the C5.0 decision tree method and also looking at the decision tree resulting from the classification
5	Accuracy Test	Carry out testing with a confusion matrix which displays accuracy tests based on training data and test data

**Testing Success Category**

In determining the categories in quality testing on systems or software, it is divided into 2 categories which can be seen in Table 2.

Table 2 Categories of Testing Success

Category	Description
Appropriate	If the quality of the software being tested is to the planning objectives and its use, then it is included in the appropriate category.
Not Appropriate	If the quality of the software being tested is not by the planning objectives and its use, then it is included in the non-conformity category.

**Testing Scenarios**

The purpose of quality testing of the system or software being built is used as a reference in carrying out quality testing of the system or software being built, as can be seen in Table 3.

Table 3 Test Scenarios

Use Case	Feature Name	Test Code	Test Case
Dataset	Import Dataset	IB-01	This feature has a function in the process of importing news datasets about politics into the system which will be stored in the database
	View Dataset	IB-02	This feature has the function of viewing datasets that have been previously imported into the system which will then display a table

Use Case	Feature Name	Test Code	Test Case
Preprocessing	Case Folding	IB-03	This feature has the function of changing all capital letters to lowercase letters where the results will be displayed by the system
	Tokenizing	IB-04	This feature has the function of breaking up the text stream into words, phrases, symbols, or other meaningful elements called tokens where the results will be displayed by the system
	Stopword Removal	IB-05	This feature has the function of deleting Stopwords such as in, to, and from where the results will be displayed by the system
	Stemming	IB-06	This feature has the function of changing each word to its basic word form where the results will be displayed by the system
Classification	News Input	IB-07	This feature has the function of inputting news into the system that will be classified which will later be stored in the database.
	Classification Results	IB-08	This feature has the function of displaying classification results to identify whether news is a hoax or valid
	See Decision Tree	IB-09	This feature has the function of displaying a decision tree on training data that has been processed using the C5.0 decision tree classification
Accuracy Test	Accuracy Test See	IB-10	This feature has the function of displaying accuracy test results on datasets that have been processed using a confusion matrix.

**Test Implementation**

At this stage, testing is carried out on the quality of the Hoax News Identification System software in the community using TF-RF and the C5.0 Decision Tree Algorithm that has been built, where this testing is carried out

concerning the designs that have been made. After that, the test results are adjusted to the objectives to be achieved from the plans that have been built which can be seen in Table 4.

Table 4. Test Implementation

No	Test Code	System Response	Expected Results	Results
1	IB-01 (Import Dataset)	Carry out the process of importing a dataset in .csv format and then saving it in the database	The system can save the dataset that has been added to the database	In accordance
2	IB-02 (See Dataset)	Displays datasets that have been added to the database and then displays them in a table	Can display datasets stored in the database	In accordance
3	IB-03 (Case Folding)	Carry out the case folding process then display the results in table form	Can carry out the case folding process and then display the case folding results stored in the database	In accordance
4	IB-04 (Tokenizing)	Carry out the tokenizing process then display the results in table form	Can carry out the tokenizing process and then display the tokenizing results stored in the database	In accordance
5	IB-05 (Stopword Removal)	Carry out the stopword removal process then display the results in table form	Can carry out the stopword removal process and then display the stopword removal results stored in the database	In accordance
6	IB-06 (Stemming)	Carry out the stemming process then display the results in table form	Can carry out the steaming process and then display the stemming results stored in the database	In accordance
7	IB-07 (News Input)	Carry out the process of inputting news that will be classified and then save it in a database accompanied by a label	Can carry out the process of inputting news and then save it in a database accompanied by a label	In accordance
8	IB-08 (Classification Results)	Carry out the process of displaying the results of the classification of political news stored in the database and then display them in the table along with the labels	Can display the results of the classification of political news stored in the database accompanied by the label	In accordance
9	IB-09 (See Decision Tree)	Displays a decision tree which is the result of using classification decision tree C5.0	Can display a decision tree on the training page where the	In accordance

No	Test Code	System Response	Expected Results	Results
			decision tree is used decision tree C5.0	
10	IB-10 (See Accuracy Test)	Displays the accuracy test of train and test data using the confusion matrix	Can display accuracy tests of training and test data with the use of the appropriate confusion matrix	In accordance

Next, the conclusion of the black box testing that has been carried out is in the table with a total of 4 functions, namely dataset, preprocessing, and accuracy testing. So the percentage of function suitability in the system can be calculated as follows:

Number of Test Codes = 10 Test Codes  
 Test Codes with Corresponding Results = 10 Test Codes  
 Test Codes with Inappropriate Results = 0 Test Codes  
 Percentage

$$\begin{aligned}
 \text{Percentage} &= \frac{(\text{number of test codes} - \text{non-conforming test codes})}{(\text{number of test codes})} \times 100\% \\
 &= \frac{(10 - 0)}{(10)} \times 100\% \\
 &= 100\%
 \end{aligned}$$

## CONCLUSION

Based on the results of testing the C5.0 decision tree algorithm and also the TF-RF weighting that has been carried out, several things have been produced. This research produces a system that can identify hoax news about politics sourced from <https://turnbackhoax.id/> and <https://www.detik.com/> automatically using a combined TF-RF method and the C5.0 algorithm. From the results of testing the system, it produces performance, namely accuracy, precision and recall on each training/test data (70/30) resulting in accuracy of 79.33%, precision 80.00%, recall 97.00%, then training data/ test (80/20) produces accuracy 79.50%, precision 81.00%, recall 95.00%, then training and test data (90/10) namely accuracy 72.00%, precision 74.00%, recall 89.00 %.

## REFERENCES

- Agustina, N., Adrian, A., & Hermawati, M. (2021). Implementasi Algoritma Naïve Bayes Classifier untuk Mendeteksi Berita Palsu pada Sosial Media. *Faktor Exacta*, 14(4), 1979–276. <https://doi.org/10.30998/faktorexacta.v14i4.11259>
- Beta, A. R., Syobah, S. N., Tahir, M., Syahab, A., & Amin, A. (2022). Literasi Digital pada Remaja dalam Upaya Menangkal Informasi Hoax Jelang Pemilu 2024. *JURNAL PPMI*, 6, 17–22.
- Deolika, A., Kusri, K., & Luthfi, E. T. (2019). Analisis Pembobotan Kata Pada Klasifikasi Text Mining. *Jurnal Teknologi Informasi*, 3(2), 179. <https://doi.org/10.36294/juriti.v3i2.1077>
- Ghaniaviyanto Ramadhan, N., Dharma Adhinata, F., Jala, A., Segara, T., Rakhmadani, P., & Informatika, F. (2022). Deteksi Berita Palsu Menggunakan Metode Random Forest dan Logistic Regression. *Jurnal Riset Komputer*, 9(2), 2407–389. <https://doi.org/10.30865/jurikom.v9i2.3979>
- Granik, M., & Mesyura, V. (2017). Fake news detection using naive Bayes classifier. *2017 IEEE 1st Ukraine Conference on Electrical and Computer Engineering, UKRCON 2017 - Proceedings*, 900–903. <https://doi.org/10.1109/UKRCON.2017.8100379>
- Handika, Y. T., Defit, S., & Nurcahyo, G. W. (2022). Text Mining Dalam Membandingkan Metode Naïve Bayes Dengan C.45 Dalam Mengidentifikasi Berita Hoax Pada Media Sosial. *Rang Teknik Journal*, 5(1), 116–123. <https://doi.org/10.31869/rtj.v5i1.2855>
- Hendriyanto, M. D., & Sari, B. N. (2022a). Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Judul Berita Hoax. *Jurnal Ilmiah Informatika*, 10(02), 80–84.
- Hendriyanto, M. D., & Sari, B. N. (2022b). Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Judul Berita Hoax. *Jurnal Ilmiah Informatika*, 10(02), 80–84.

- Irena, B., & Erwin Budi Setiawan. (2020). Fake News (Hoax) Identification on Social Media Twitter using Decision Tree C4.5 Method. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(4), 711–716. <https://doi.org/10.29207/resti.v4i4.2125>
- Meo, M. O. (2017). Penerapan Algoritme Pohon Keputusan C5.0 Untuk Klasifikasi Lahan Gambut Yang Terbakar Di Kabupaten Ogan Komering Ilir. *Jurnal Teknologi Terpadu*, 3(1), 17–23.
- Prayoga, A. Y., Hadiana, A. I., & Umbara, F. R. (2021a). Deteksi Hoax Pada Berita Online Bahasa Inggris Menggunakan Bernoulli Naive Bayes Dengan Ekstraksi Fitur TF-IDF. *Sosial Teknik*, 2(3), 1809–1823.
- Prayoga, A. Y., Hadiana, A. I., & Umbara, F. R. (2021b). Deteksi Hoax Pada Berita Online Bahasa Inggris Menggunakan Bernoulli Naive Bayes Dengan Ekstraksi Fitur TF-IDF. *Sosial Teknik*, 2(3), 1809–1823.
- Sains, S., Kelurahan, D., Hilir, S., Leman, D., Rahman, M., & Ramadhan, M. H. (2022a). Analisis Metode C5 . 0 Untuk Penyediaan Lapangan Pekerjaan. *Sains Dan Teknologi Informasi*, 8(2), 112–121. <https://doi.org/10.33372/stn.v8i2.906>
- Sains, S., Kelurahan, D., Hilir, S., Leman, D., Rahman, M., & Ramadhan, M. H. (2022b). Analisis Metode C5 . 0 Untuk Penyediaan Lapangan Pekerjaan. *Sains Dan Teknologi Informasi*, 8(2), 112–121. <https://doi.org/10.33372/stn.v8i2.906>
- Utami, M. P., Nurhayati, O. D., & Warsito, B. (2020). Hoax Information Detection System Using Apriori Algorithm and Random Forest Algorithm in Twitter. *6th International Conference on Interactive Digital Media, ICIDM 2020, Icidm*. <https://doi.org/10.1109/ICIDM51048.2020.9339648>
- Utomo, A., Sutanto, Y., Tiningrum, E., & Susilowati, E. M. (2020). *Pengujian Aplikasi Transaksi Perdagangan Menggunakan Black Box Testing Boundary Value Analysis*. 04, 133–140.
- Wati, R. (2020). Penerapan Algoritma Naive Bayes Dan Particle Swarm Optimization Untuk Klasifikasi Berita Hoax Pada Media Sosial. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 5(2), 159–164. <https://doi.org/10.33480/jitk.v5i2.1034>
- Willy, Setiawan, E. B., & Nugraha, F. N. (2019). Implementation of Decision Tree C4.5 for Big Five Personality Predictions with TF-RF and TF-CHI2 on Social Media Twitter. *2019 International Conference on Computer, Control, Informatics and Its Applications: Emerging Trends in Big Data and Artificial Intelligence, IC3INA 2019*, 114–119. <https://doi.org/10.1109/IC3INA48034.2019.8949601>