

unplug Implementation of ID3 algorithm classification using web-based weka

by Mochamad Alfian Rosid

Submission date: 09-Jan-2024 12:20PM (UTC+0700)

Submission ID: 2268248231

File name: Fitrani_2019_J._Phys._Conf._Ser._1381_012036.pdf (873.4K)

Word count: 2478

Character count: 12519

PAPER · OPEN ACCESS

Implementation of ID3 algorithm classification using web-based weka

6

1
To cite this article: A S Fitriani *et al* 2019 *J. Phys.: Conf. Ser.* **1381** 012036

1
View the [article online](#) for updates and enhancements.

You may also like

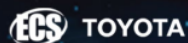
- [Analysis of data mining classification by](#)

2 [Comparison of C4.5 and ID algorithms](#)
R. Sudrajat, I. Irianingsih and D. Krisnawan

- [Comprehensive Gas Characterization of a z = 2.5 Protocluster: A Cluster Core Caught in the Beginning of Virialization?](#)
Jaclyn B. Champagne, Caitlin M. Casey, Jorge A. Zavala et al.

- [Establishment and Improvement of Financial Decision Support System Using Artificial Intelligence and Big Data](#)
Cong Si and Wenquan Shi

ECS Toyota Young Investigator Fellowship



For young professionals and scholars pursuing research in batteries, fuel cells and hydrogen, and future sustainable technologies.

At least one \$50,000 fellowship is available annually.
More than \$1.4 million awarded since 2015!



Application deadline: January 31, 2023

Learn more. Apply today!

5 Implementation of ID3 algorithm classification using web-based weka

3
A S Fitriani*, M A Rosid, Y Findawati, Y Rahmawati, A K Anam

Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Jl. Raya Gelam 250
Candi Sidoarjo, Jawa Timur, Indonesia

asfjim@umsida.ac.id

Abstract. The Bangil District Court is an IB class court that handles a large number of case cases. Every year more and more case cases are included in the Bangil District Court, but not all case cases are in a mutation status. By using classification techniques that can process large amounts of data to find patterns that occur in case data. Data processing is used to predict case minutation with the decision tree method using ID3 algorithm. Case data has 8 attributes and has been classified into 6 parts, namely division based on Case Type, Register, Case Classification, Length of Process, Public Prosecutor and Decision with a goal of Mutation Status. Weka 3.6 is an API that is used to build rules / rule bases. The rule that was formed was then implemented in the making of a case status prediction application in the web-based Bangil District Court.

1. Introduction

The Bangil District Court which is an IB class court that along with the increasing number of case cases in the Bangil District Court, the number of incoming case data automatically increases, but not all case cases are in a mutation status. This case data will be processed using data mining techniques that will generate new knowledge and information from patterns or models formed on large data.

By using classification techniques that can process large amounts of data[1][2][3] where it will be explained about the model produced by the data that has a data class on case data in Pengadilan Negeri Bangil. Data processing is used to predict case minutation with the decision tree method using ID3 algorithm.

Case data taken is data from the website of the Case Search Information System (SIPP) in the Bangil District Court. Data will be processed using the iterative classification method dichotomiser 3 (Id3) to find out predictions of case mutations from 2011 to 2015 in the Bangil District Court.


The impact of this application is that the IT Staff of the Bangil District Court is easier and more efficient in terms of completing the case data prediction process to find out the status of the mutation with ID3 algorithm.

2. Theoretical foundation

2.1 Definition of Case Search Information System

4
The Case Search Information System (SIPP) is an official web-based application made by the Supreme Court of the Republic of Indonesia (MARI) to support and facilitate justice seekers in monitoring the process of handling criminal cases[4].

This application is intended for administration and tracking of first-degree court case data. Case Phase, Case Status, Case Costs, Session Schedule, Case Statistics, and History, including information services contained in this application and complete information, has been recorded.

1
 Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

All **information** can be accessed by the general public for free and realtime. This SIPP application allows access to case files. The community can directly access and search for the subject matter [5]

2.2 Definition of Data Mining

Data Mining is a process of collecting data that is processed with various methods. Another term for data mining is knowledge-discovery in a database (KDD). The purpose of data mining is to utilize data and process it to get new and useful information.

2.3 Group Data Mining

There are several types of data mining groupings, according to [6]:

1. Description, to describe the pattern in data.
2. Estimates, the goal is more numerical than category, and this estimate is almost similar to classification.
3. Prediction, the result of a prediction is to show something that hasn't happened yet
4. Classification, the purpose of this classification is categorical
5. Clustering, clustering is more towards grouping records or cases that have similar data
6. Association, to identify the relationship between events that occur at one time.

2.4 Stages of Data Mining

The data mining stage is interactive, and the stages of data mining are divided into several stages including. [6]

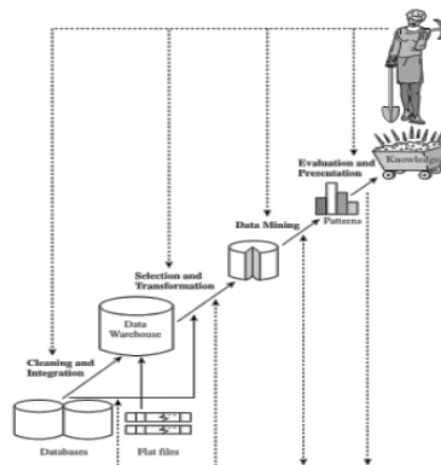


Figure 1. Stages of Data Mining

1. Cleaning (Cleaning)
2. Is the process of removing/cleaning inconsistent data.
3. Integration (integration)
4. Is the merging of data from various databases into one new database
5. Selection
6. Data selection (data selection), not all data contained in the database is used; the data taken is only suitable for analysis.
7. Transformation, data will be converted into a format that is compatible with the data mining process. **3**
8. Data mining process, is the most important process in the method used to find knowledge.
9. Evaluate patterns (pattern evaluation), Identify a unique pattern and draw it into the knowledge base

10. Presentation of knowledge (knowledge presentation), is a presentation of knowledge from the method used to obtain knowledge obtained by the user.

2.5 Decision Tree Method

Decision Tree is an approach that is widely used to solve clarification problems, and this method is used to estimate discrete values and target functions. The Decision Tree is a set of IF-THEN rules that each line or path is connected to the rule. In the decision, the tree can be used to map decision [7].

2.5.1 ID3 algorithm

ID3 algorithm is one method in the decision tree model, where estimating the discrete value of a function from a decision tree produces an if-then rule so that it is easier to understand. The following is an example of the Iterative Dichonomiser 3 or ID3 algorithm system [8][9].

2.5.2 Entropy

Entropy is a measure of data that has a measure of information theory that can determine the characteristics of data impurity and feature equations of members of an attribute[1][8]. To calculate the entropy value using the formula:

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (1)$$

Information:

S: a set of cases

n: number of S partitions

Pi: the proportion of Si to S

2.5.3 Gain

After the entropy value is known, the next step is to find the gain value of an attribute in classifying the data[1][10].

To calculate the Gain value using the formula:

$$Gain(S,A)=Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i) \quad (2)$$

Information:

S: a set of cases

A: Attribute

n: number of partition attributes A

| Si |: number of cases on a partition to i

| S |: number of cases in S

3. System Design

The design of the application of data mining prediction on minutation cases can be divided into the following stages, namely:

1. Classification Diagram
2. Data Attributes
3. Flowchart
4. Designing the User Interface

3.1 Classification Diagram

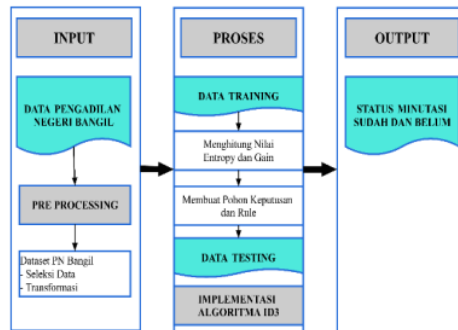


Figure 2. Classification Diagram

3.2 Flowchart

The following is a description of the flowchart classification method with algorithm ID3 to predict case status minutation, namely:

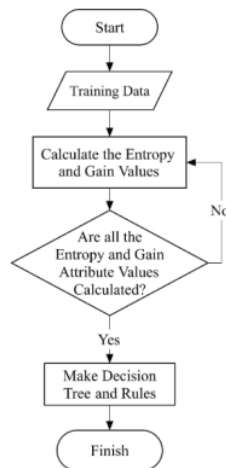


Figure 3. Flowchart Data Training

The description of figure 3 is:

1. Enter training data
2. Calculate the entropy and gain values
3. Create a root node (root node) based on the highest Gain value.
4. Calculate the Entropy value based on the attribute on the selected Gain.
5. Calculate all entropy values until you have one answer or leaf node if you have not found one answer, it needs to be recalculated by choosing the largest gain value remaining. After all, are counted, create a branch node (internal node) based on the results of the calculation of entropy.
6. After Gain and Entropy have calculated all, make a decision tree and Rule from the decision tree that has been made.

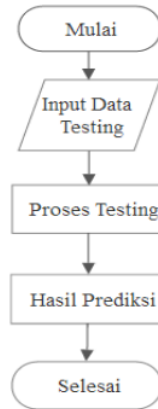


Figure 4. Flowchart Data Testing

The information from figure 4 is:

- Input testing data
- The testing process is based on the decision tree from the training data
- Get the results of testing data

4. Discussion

4.1 Input

Data obtained from the Bangil District Court totaled 998 case data. The data will be tested using the Decision Tree method with algorithm ID3 using WEKA tools and website[2]. Data collection was 65% of 649 training data and 35% of 349 testing data. Following are some examples of case data, namely:

Table 1. Dataset

No	Terdakwa/Pemohon	Jenis Perkara	Register	Klasifikasi Perkara	Lama Proses	Penuntut Umum	Putusan	Status Minutasi
1	IRA SUGIARTI	Pdt.G	A	Ganti Rugi	Sangat Lama	TD	Dikabulkan	Behum
2	PT. BPR PANDAAN ARTA JAYA	Pdt.G	A	Wanprestasi	Sangat Lama	TD	Dikabulkan	Behum
3	WARTI DAN PINI	Pdt.G	A	Wanprestasi	Sangat Lama	TD	Dicabut	Sudah
4	M. SYAIFUDDIN NUR	Pid.B	A	Pembunuhan	Lama	TD	Penjara	Behum
5	YOYOK SAPTOADI JOKO SUPRIJOKO	Pid.B	A	Penggelapan	Cepat	TD	Penjara	Behum
6	DAIM	Pid.B	A	Narkotika	Lama	TD	Penjara	Behum
7	FATHUR ROHMAN	Pid.B	A	Penggelapan	Lama	TD	Penjara	Behum
8	RAKHMAT HIDAYAT	Pid.B	A	Narkotika	Lama	TD	Penjara	Behum
9	H. NUR HASAN	Pid.B	A	Pencurian	Lama	TD	Penjara	Behum
10	INUNG	Pid.B	A	Narkotika	Sangat Lama	TD	Penjara	Behum
11	IDA BAGUS KOKO	Pid.B	A	Lain-Lain	Sangat Lama	TD	Bebas	Behum
12	IRA SUGIARTI	Pdt.G	A	Ganti Rugi	Sangat Lama	TD	TDD	Behum
13	Disamakan	Pdt.G	B	Penceraian	Cepat	D	Dikabulkan	Sudah
14	Disamakan	Pdt.G	B	Penceraian	Sangat Cepat	D	Dikabulkan	Sudah
15	Disamakan	Pdt.G	B	Penceraian	Cepat	D	Dikabulkan	Sudah
16	PT. BPR GUNUNG ADI DANA	Pdt.G	B	Lain-Lain	Sangat Lama	TD	Dikabulkan	Behum
17	PT. CIPTA MAKARYA INDAH	Pdt.G	B	Lain-Lain	Lama	TD	Dicabut	Sudah
18	GINTEN	Pdt.G	B	Lain-Lain	Sangat Lama	TD	Dikabulkan	Behum
19	Disamakan	Pdt.G	B	Penceraian	Sangat Cepat	D	Dikabulkan	Sudah
20	LILA HERDIANTI	Pdt.G	B	Ganti Rugi	Cepat	TD	Dikabulkan	Sudah

Table 2. Training Data

No	Terdakwa/Pemohon	Jenis Perkara	Register	Klasifikasi Perkara	Lama Proses	Penuntut Umum	Putusan	Status Minutasi
1	IRA SUGIARTI	Pdt.G	A	Ganti Rugi	Sangat Lama	TD	Dikabulkan	Behm
2	PT. BPR PANDAAN ARTA JAYA	Pdt.G	A	Wanprestasi	Sangat Lama	TD	Dikabulkan	Behm
3	YOYOK SAPTOADI JOKO SUPRIOKO	Pid.B	A	Penggelapan	Cepat	TD	Penjara	Behm
4	DAIM	Pid.B	A	Narkotika	Lama	TD	Penjara	Behm
5	FATHUR ROHMAN	Pid.B	A	Penggelapan	Lama	TD	Penjara	Behm
6	RAKHMAT HIDAYAT	Pid.B	A	Narkotika	Lama	TD	Penjara	Behm
7	H. NUR HASAN	Pid.B	A	Pencurian	Lama	TD	Penjara	Behm
8	INUNG	Pid.B	A	Narkotika	Sangat Lama	TD	Penjara	Behm
9	IDA BAGUS KOKO	Pid.B	A	Lain-Lain	Sangat Lama	TD	Bebas	Behm
10	IRA SUGIARTI	Pdt.G	A	Ganti Rugi	Sangat Lama	TD	TDD	Behm
11	Disamakan	Pdt.G	B	Penceraian	Cepat	D	Dikabulkan	Sudah
12	Disamakan	Pdt.G	B	Penceraian	Sangat Cepat	D	Dikabulkan	Sudah
13	Disamakan	Pdt.G	B	Penceraian	Cepat	D	Dikabulkan	Sudah
14	PT. BPR GUNUNG ADI DANA	Pdt.G	B	Lain-Lain	Sangat Lama	TD	Dikabulkan	Behm
15	PT. CIPTA MAKARYA INDAH	Pdt.G	B	Lain-Lain	Lama	TD	Dicabut	Sudah
16	GINTEN	Pdt.P	B	Lain-Lain	Sangat Lama	TD	Dikabulkan	Behm
17	Disamakan	Pdt.G	B	Penceraian	Sangat Cepat	D	Dikabulkan	Sudah
18	LILA HERDIANTI	Pdt.G	B	Ganti Rugi	Cepat	TD	Dikabulkan	Sudah
19	LILA HERDIANTI	Pdt.G	B	Ganti Rugi	Lama	TD	Dikabulkan	Sudah
20	NINDIA AVISTA CITRAYANTI	Pdt.P	B	PAKT	Sangat Cepat	TAO	Dikabulkan	Sudah

Table 3. Testing Data

No	Terdakwa/Pemohon	Jenis Perkara	Register	Klasifikasi Perkara	Lama Proses	Penuntut Umum	Putusan	Status Minutasi
1	WARTI DAN PINI	Pdt.G	A	Wanprestasi	Sangat Lama	TD	Dicabut	Sudah
2	M. SYAIFUDDIN NUR	Pid.B	A	Pembunuhan	Lama	TD	Penjara	Behm
3	BAMBANG APRIYANTO	Pid.B	B	Penggelapan	Cepat	TD	Penjara	Sudah
4	KHOLIK	Pdt.G	B	Lain-Lain	Lama	TD	Dikabulkan	Sudah
5	SUNARCO Als REJES Bin DASIRUN	Pid.B	B	Penggelapan	Cepat	TD	Penjara	Sudah
6	H.HILMI	Pdt.P	B	PAKT	Sangat Cepat	TAO	Dikabulkan	Sudah
7	DIANTO	Pdt.P	B	PAKT	Sangat Cepat	TAO	Dikabulkan	Sudah
8	NURJANAH	Pdt.P	B	WDJ	Sangat Cepat	TAO	Dikabulkan	Sudah
9	MAWARDI	Pdt.P	B	WDJ	Sangat Cepat	TAO	Dikabulkan	Sudah
10	Disamakan	Pdt.G	B	Penceraian	Sangat Cepat	D	Dikabulkan	Sudah
11	TRI SETYO JANUARIFIN	Pdt.P	B	PAKT	Sangat Cepat	TAO	Dikabulkan	Sudah
12	SRI RAHAYU YULIARTI	Pid.B	B	Pemalsuan	Cepat	TD	Penjara	Behm
13	Terdakwa	Pid.B	B	Lain-Lain	Cepat	TD	Penjara	Sudah
14	NURIYANTO	Pdt.P	B	PAKT	Sangat Cepat	TAO	Dikabulkan	Sudah
15	SAHUR BIN MARSAN	Pid.B	B	Penganiayaan	Sangat Cepat	TD	Penjara	Sudah
16	RUHANAH	Pdt.P	B	PAKT	Sangat Cepat	TAO	Dikabulkan	Sudah
17	YAYUK SUPRPTI BROWNE	Pdt.P	B	PAKT	Cepat	TAO	Dikabulkan	Sudah
18	LOLOK WIJAYANTI	Pdt.G	B	Lain-Lain	Sangat Lama	TD	Dikabulkan	Sudah
19	MITA FITRIA	Pdt.P	B	PAKT	Sangat Cepat	TAO	Dikabulkan	Sudah
20	USMANTO BIN NGADIARTO	Pid.B	B	Pencurian	Cepat	TD	Penjara	Sudah

4.1.1 Data

The dataset used for the calculation process is 998 case data. The dataset is divided into two parts, namely training data and testing data. Training data is used to process a calculation by forming a classifier model. While data testing is used to measure the extent to which the classifier successfully classifies correctly.

4.1.2 Pre Processing Data

The 1,000 datasets, which were the initial data of this study experienced pre-processing to 998 data. Where the pre-processing techniques in preparing data are discarding duplicate data, discarding data inconsistencies, and correcting data errors.

4.2 Proses

The process used at this stage is to find out the results of the ID3 classification calculation on the Website.

4.2.1 Website

1. After that, take the dataset with the extension .csv / .arff

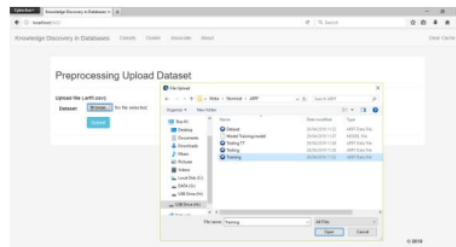


Figure 5. Display of Dataset Files

2. Select the classify method, select the ID3 algorithm and then select the dataset and Testing data and then select submit

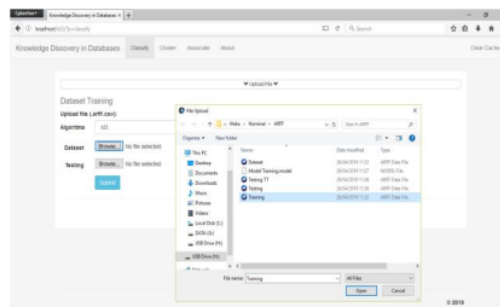


Figure 6. Classify Display

3. Wait for a while until the classification results appear

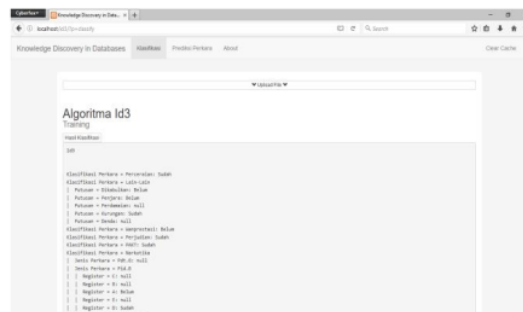


Figure 7. Display of ID3 Classification Results

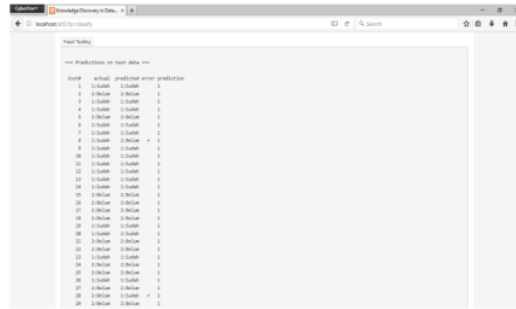


Figure 8. Display of Testing Results

4. Display prediction menu.

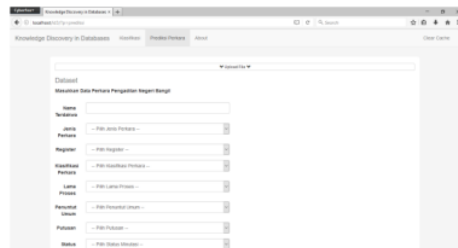


Figure 9. Display Prediction

5. Wait for a while until the prediction results appear



Figure 10. Display of Prediction Results

4.3 Output

4.3.1 Classification Results

1. Predicted results on testing with a total testing data of 35% of a total dataset of 998.

```

===== Predictions on test data =====
inst#  actual  predicted error prediction
1  2:Sudah  1:Belum  + 1
2  1:Belum  ?      ?
3  2:Sudah  ?      ?
4  2:Sudah  2:Sudah  1
5  2:Sudah  ?      ?
6  2:Sudah  2:Sudah  1
7  2:Sudah  2:Sudah  1
8  2:Sudah  2:Sudah  1
9  2:Sudah  2:Sudah  1
10 2:Sudah  2:Sudah  1
... ..
    
```

2. Prediction results on testing with single data testing.

```

===== Predictions on test data =====
inst#  actual  predicted error prediction
1  1:Belum  2:Sudah  + 1
    
```

4.3.2 Percentage Results

Percentage of processes to determine accuracy. The following are the results of the percentage test accuracy Table 4 :

Table 4. Percentage Results

Data Testing	Result			
	True	%	False	%
988	326	93.41 %	23	6.59 %

5. Conclusion

Based on the results of the study, it can be concluded that the algorithm ID3 can be considered to be used in the case data prediction process to determine the mutation status. This is supported by the results of testing the system using algorithm ID3 in the Bangil District Court, which has an average accuracy of 93.41% and has an error rate of 6.59%.

Suggestions that can be given for this research are expected to be able to use other methods. So, can know the comparison between the percentage results in the calculation process according to the method used.

6. References

- [1] Gupta B, Rawat A, Jain A, Arora A and Dhama N 2017 Analysis of Various Decision Tree Algorithms for Classification in Data Mining *Int. J. Comput. Appl.* **163** 15–9
- [2] Guo H, Hong N, Shen Z, Duan W and Zhang Z 2018 Application of Classification Technique of Data Mining for Employee Management System **1** 12–22
- [3] Sharma P, Singh D and Singh A 2015 Classification algorithms on a large continuous random dataset using rapid miner tool *2nd Int. Conf. Electron. Commun. Syst. ICECS 2015* 704–9
- [4] Putra A 2015 PEMBAHARUAN SISTEM PERADILAN PIDANA MELALUI PENATAAN ADMINISTRASI PERADILAN **IV** 73–9
- [5] Nisa K and Hany Fanida E 2000 PENERAPAN SISTEM INFORMASI ADMINISTRASI PERKARA PEGADILAN AGAMA (SIADPAPLus) MELALUI APLIKASI AUDIO TO TEXT RECORDING DI PENGADILAN AGAMA KABUPATEN MALANG Khoirun Nisa *Unasa* **1204067400** 1–11
- [6] Ridwan M, Suyono H and Sarosa M 2013 Jurnal Pak Agung **7** 59–64
- [7] B L K 2016 Algorithm Selection for Combinatorial Search Problems: A Survey **10101** 149–90
- [8] Wang Y Y, Li Y Bin and Rong X W 2017 Improvement of ID3 algorithm based on simplified information entropy and coordination degree *Proc. - 2017 Chinese Autom. Congr. CAC 2017* **2017-Janua** 1526–30
- [9] Jassim K S and Saleh H M 2018 Diabetes Classification Using ID3 and Naïve Bayes Algorithms **12**
- [10] Kraidech S and Jearanaitanakij K 2018 Improving ID3 Algorithm by Combining Values from Equally Important Attributes *ICSEC 2017 - 21st Int. Comput. Sci. Eng. Conf. 2017, Proceeding* **6** 102–5

unplug Implementation of ID3 algorithm classification using web-based weka

ORIGINALITY REPORT

10%

SIMILARITY INDEX

8%

INTERNET SOURCES

9%

PUBLICATIONS

3%

STUDENT PAPERS

PRIMARY SOURCES

1

elar.urfu.ru

Internet Source

3%

2

I M K Karo, M Y Fajari, N U Fadhilah, W Y Wardani. "Benchmarking Naïve Bayes and ID3 Algorithm for Prediction Student Scholarship", IOP Conference Series: Materials Science and Engineering, 2022

Publication

3%

3

Mochamad Alfian Rosid, Arif Senja Fitriani, Ika Ratna Indra Astutik, Nasrudin Iqrok Mulloh, Haris Ahmad Gozali. "Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi", IOP Conference Series: Materials Science and Engineering, 2020

Publication

1%

4

ojs.unik-kediri.ac.id

Internet Source

1%

5

Jin Cui. "Using Machine Learning Algorithms to Predict Efficiency of US's States Mask Policy"

1%

During COVID-19", Journal of Physics: Conference Series, 2021

Publication

6

www.science.gov

Internet Source

1 %

Exclude quotes Off

Exclude matches < 1%

Exclude bibliography On