

*Reprinted From*

---

ISSN: 1742-6596 EISSN: 1742-6596

## **Journal of Physics: Conference Series**

Volume 801, Number 1

<http://iopscience.iop.org>

---

---

# **CORRUPTION CASES MAPPING BASED ON INDONESIA'S CORRUPTION PERCEPTION INDEX**

<sup>1</sup>Noerlina, <sup>2</sup>L A Wulandari, <sup>3</sup>Sasmoko, <sup>4</sup>Andi M Muqsith and <sup>5</sup>M Alamsyah

<sup>1</sup>School of Information System, Bina Nusantara University, Jakarta, Indonesia, 11480

<sup>2,4,5</sup>School of Computer Science, Bina Nusantara University, Jakarta, Indonesia, 11480

<sup>3</sup>Research Interest Group in Educational Technology, Bina Nusantara University, Jakarta, Indonesia, 11480

<sup>3</sup>Faculty of Humanities, Bina Nusantara University, Jakarta, Indonesia, 11480

[URL: http://iopscience.iop.org/article/10.1088/1742-6596/801/1/012019](http://iopscience.iop.org/article/10.1088/1742-6596/801/1/012019)



IOPscience

# Document details

< Back to results | < Previous 16 of 24 Next >

Save to Mendeley Download Print E-mail Save to PDF Save to list More... >

View at Publisher

Journal of Physics: Conference Series  
Volume 801, Issue 1, 27 March 2017, Article number 012019  
1st International Conference on Computing and Applied Informatics, ICCAI 2016; Hotel GranDhika Setiabudi MedanMedan; Indonesia; 14 December 2016 through 15 December 2016; Code 127044

## Corruption Cases Mapping Based on Indonesia's Corruption Perception Index (Conference Paper) [\(Open Access\)](#)

Noerlina<sup>a</sup>, Wulandhari, L.A.<sup>b</sup>, Sasmoko<sup>c,d</sup>, Muqsith, A.M.<sup>b</sup>, Alamsyah, M.<sup>b</sup>

<sup>a</sup>School of Information System, Bina Nusantara University, Indonesia

<sup>b</sup>School of Computer Science, Bina Nusantara University, Indonesia

<sup>c</sup>Faculty of Humanities, Bina Nusantara University, Indonesia

View additional affiliations

### Abstract

View references (15)

Government plays an important role in nation economic growth. Nevertheless, there are still many occurrences of government officers abusing their offices to do an act of corruption. In this order, the central government should pay attention to every area in the nation to avoid corruption case. Meanwhile, the news media always constantly preach about corruption case, this makes the news media relevant for being one of the sources of measurement of corruption perception index (CPI). It is required to map the corruption case in Indonesia so the central government can pay attention to every region in Indonesia. To develop the mapping system, researchers use Naïve Bayes Classifier to classify which news articles talk about corruption and which news articles are not, before implementing a Naïve Bayes Classifier there are some text processing such as tokenizing, stopping, and stemming. © Published under licence by IOP Publishing Ltd.

### Indexed keywords

Engineering controlled terms: Economics Mapping Sodium Text processing

Engineering uncontrolled terms: Bayes Classifier Central government Corruption perception indices Economic growths Mapping systems News articles News media Tokenizing

Engineering main heading: Crime

ISSN: 17426588  
Source Type: Conference Proceeding  
Original language: English

DOI: 10.1088/1742-6596/801/1/012019  
Document Type: Conference Paper  
Sponsors:  
Publisher: Institute of Physics Publishing

### References (15)

View in search results format >

All | Save to Mendeley Print E-mail Save to PDF Create bibliography

### Metrics View all metrics >

1 Citation in Scopus  
82nd Percentile

4.63 Field-Weighted  
Citation Impact



### PlumX Metrics

Usage, Captures, Mentions, Social Media and Citations beyond Scopus.

### Cited by 1 document

Corruption system development based on Indonesia's corruption perception index  
Noerlina, Wulandhari, L.A., Arifin, Y. (2018) *ICIC Express Letters, Part B: Applications*

View details of this citation

Inform me when this document is cited in Scopus:

[Set citation alert >](#)

[Set citation feed >](#)

### Related documents

Latent Dirichlet Allocation (LDA) Model and kNN Algorithm to Classify Research Project Selection  
Saf'ie, M.A., Utami, E., Fatta, H.A. (2018) *IOP Conference Series: Materials Science and Engineering*

How Computer Science Risks to Lose Its Innocence, and Should Attempt to Take Responsibility  
Aberer, K. (2017) *Proceedings - International Conference on Distributed Computing Systems*

- 1 Ghosh, R.N., Siddique, M.A.B.  
**Corruption, good governance and economic development: Contemporary analysis and case studies**

(2014) *Corruption, Good Governance and Economic Development: Contemporary Analysis and Case Studies*, pp. 1-266. Cited 2 times.

<http://www.worldscientific.com/worldscibooks/10.1142/9207#t=oc>

ISBN: 978-981461259-3; 978-981461258-6

doi: 10.1142/9789814612593

[View at Publisher](#)

A non deterministic Indonesian stemmer

Purwarianti, A.  
(2011) *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, ICEEI 2011*

[View all related documents based on references](#)

[Find more related documents in Scopus based on:](#)

[Authors >](#) [Keywords >](#)

- 2 Huang, J., Lu, J., Ling, C.  
(2003) *Comparing Naive Bayes, Decision Trees, and SVM with AUC and Accuracy.* Cited 2 times.

- 3 Jang, S., Woo, W.  
5W1H: Unified User-Centric Context  
(2005) *Proceedings of the 7th International Conference on Ubiquitous Computing.*

- 4 Malik, D.  
(2001) *Dari Konstruksi Ke Dekonstruksi: Refleksi Atas Pemberitaan Televisi Kita.*

- 5 Melgar, N., Rossi, M., Smith, T.W.  
**The perception of corruption**  
(2010) *International Journal of Public Opinion Research*, 22 (1), pp. 120-131. Cited 38 times.  
doi: 10.1093/ijpor/edp058

[View at Publisher](#)

- 6 Murphy, K.P.  
(2006) *Naive Bayes Classifiers Generative Classifiers*, 4701, pp. 1-8.

- 7 Olteanu, A., Peshterliev, S., Liu, X., Aberer, K.  
**Web credibility: Features exploration and credibility prediction**  
(2013) *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7814 LNCS, pp. 557-568. Cited 23 times.  
ISBN: 978-364236972-8  
doi: 10.1007/978-3-642-36973-5\_47

[View at Publisher](#)

- 8 Persson, A., Rothstein, B., Teorell, J.  
**Why anticorruption reforms fail-systemic corruption as a collective action problem**  
(2013) *Governance*, 26 (3), pp. 449-471. Cited 122 times.  
doi: 10.1111/j.1468-0491.2012.01604.x

[View at Publisher](#)

- 
- 9 Rahimi, R., Shakeri, H.  
Reflection on Judicial System's Corruption and Offering Solutions to Promote its Safety  
(2016) *Journal of Politics and Law*, 9 (9), p. 187.
- 
- 10 Rohwer, A.  
Measuring corruption: A comparison between the transparency international's corruption perceptions index and the world bank's worldwide governance indicators  
(2009) *CESifo DICE Report*, 7 (3), pp. 42-52. Cited 19 times.
- 
- 11 Schrenk, M.  
(2007) *Webbots, Spiders, and Screen Scrapers*.
- 
- 12 Tala, F.  
(2003) *A Study of Stemming Effects on Information Retrieval in Bahasa Indonesia*. Cited 46 times.
- 
- 13 Tated, R.R., Ghonge, M.M.  
A survey on text mining - Techniques and application  
(2015) *International Journal of Research in Advent Technology*, 1, pp. 380-385. Cited 2 times.
- 
- 14 Uca, N., Ince, H., Sumen, H.  
The Mediator Effect of Logistics Performance Index on the Relation between Corruption Perception Index and Foreign Trade Volume  
(2016) *European Scientific Journal*, 12 (25), pp. 1857-7881.
- 
- 15 Yang, J.M., Cai, R., Wang, C., Huang, H., Zhang, L., Ma, W.Y.  
Incorporating site-level knowledge for incremental crawling of web forums: A list-wise strategy  
(2009) *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1375-1383. Cited 13 times.  
ISBN: 978-160558495-9  
doi: 10.1145/1557019.1557166  
  
View at Publisher
- 

© Copyright 2018 Elsevier B.V., All rights reserved.

[< Back to results](#) | [< Previous](#) 16 of 24 [Next >](#)

[^ Top of page](#)

## About Scopus

[What is Scopus](#)  
[Content coverage](#)  
[Scopus blog](#)  
[Scopus API](#)

## Language

[日本語に切り替える](#)  
[切换到简体中文](#)  
[切换到繁體中文](#)  
[Русский язык](#)

## Customer Service

[Help](#)  
[Contact us](#)

ELSEVIER

[Terms and conditions](#) [Privacy policy](#)

Copyright © 2018 Elsevier B.V. All rights reserved. Scopus® is a registered trademark of Elsevier B.V.

Cookies are set by this site. To decline them or learn more, visit our [Cookies page](#).

 RELX Group™

# International Conference on Computing and Applied Informatics 2016

---

The Faculty of Computer Science and Information Technology, University of Sumatera Utara is pleased to announce the 1st International Conference on Computing and Applied Informatics (ICCAI 2016) which will be held in Medan, Indonesia on December 14-15, 2016. This event provides an excellent platform for knowledge exchange between academicians, researchers, scientists, engineers and stakeholders working in the areas related to computing and applied informatics fields. The conference also offers the participants opportunities to network, collaborate, as well as to share research findings and experiences.

## Keynote Speaker

1. Prof. Hiromi Homma (Emeritus Professor from Toyohashi University of Technology, Japan)
2. Prof. Madya Dr. Rahmita Wirza O. K. Rahmat (Distinguished Professor from Universiti Putra Malaysia)

## Honorary Chair

Prof. Dr. Runtung Sitepu, S.H., M.Hum (Rector of University of Sumatera Utara)

## International Board

Prof. Hiromi Homma (Toyohashi University of Technology, Japan)  
Prof. Dr. Masashi Daimaruya (Muroran Institute of Technology, Japan)  
Prof. Dr. Rahmat Budiarto (Al Baha University, Saudi Arabia)  
Muhammad Fermi Pasha, B.Comp.Sc., M.Sc., Ph.D. (Monash University, Malaysia Campus)

## Steering Committee

Prof. Dr. Hideki Kawai (Muroran Institute of Technology, Japan)  
Prof. Zainal A. Hasibuan, Ir., MLS, PhD (University of Indonesia, Indonesia)  
Prof. Dr. Opim Salim Sitompul (University of Sumatera Utara, Indonesia)  
Drs. Mahyuddin K. M. Nasution, M.IT, PhD (University of Sumatera Utara, Indonesia)  
Prof. Dr. Herman Mawengkang (University of Sumatera Utara, Indonesia)  
Prof. Dr. Saib Suwilo (University of Sumatera Utara, Indonesia)  
Dr. Ir. Rinaldi Munir, MT (Bandung Institute of Technology, Indonesia)  
Prof. Dr. rer.nat. Dedi Rosadi, M.Sc. (University of Gadjah Mada, Indonesia)  
Benny B. Nasution, Ph.D (Politeknik Negeri Medan, Indonesia)  
Dr. Fitri Arnia, M.Eng.Sc (Syiah Kuala University, Indonesia)

## Organizing Committee

### Chair

Dr. Eng. Himsar Ambarita

### Co. Chair

Romi Fadilah Rahmat, B.Comp.Sc., M.Sc  
Suherman, PhD

### Members

Dr. Erna Budhiarti Nababan, M.IT

Emerson Sinulingga, PhD

Dr. Fahmi F.

Mohammad Fadly Syah Putra, M.Sc

Muhammad Anggia Muchtar, ST, MMIT

Dani Gunawan, S.T, M.T

Taufiq Bin Nur, ST, M.Eng.Sc

## Corruption Cases Mapping Based on Indonesia's Corruption Perception Index

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2017 J. Phys.: Conf. Ser. 801 012019

(<http://iopscience.iop.org/1742-6596/801/1/012019>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 114.124.31.89

This content was downloaded on 29/03/2017 at 04:01

Please note that [terms and conditions apply](#).

You may also be interested in:

[Corruption Early Prevention: Decision Support System for President of the Republic of Indonesia](#)  
Sasmoko, S A Widhoyoko, S Ariyanto et al.

[Adaptive classifier for steel strip surface defects](#)  
Mingming Jiang, Guangyao Li, Li Xie et al.

[CargoCBM – Feature Generation and Classification for a Condition Monitoring System for Freight Wagons](#)  
C Gericke and M Hecht

[Classification of autism spectrum disorder using supervised learning of brain connectivity measures extracted from synchronostates](#)  
Wasifa Jamal, Saptarshi Das, Ioana-Anastasia Oprescu et al.

[NLP with dynamic classification improves P300 speller accuracy and bit rate](#)  
William Speier, Corey Arnold, Jessica Lu et al.

[Identifying early stage precipitation in large-scale atomistic simulations of superalloys](#)  
Eric Schmidt and Paul D Bristowe

[Using complex networks for text classification: Discriminating informative and imaginative documents](#)  
Henrique F. de Arruda, Luciano da F. Costa and Diego R. Amancio

[Decoding vowels and consonants in spoken and imagined words using electrocorticographic signals in humans](#)  
Xiaomei Pei, Dennis L Barbour, Eric C Leuthardt et al.



# Corruption Cases Mapping Based on Indonesia's Corruption Perception Index

Noerlina<sup>1</sup>, L A Wulandhari<sup>2</sup>, Sasmoko<sup>3,4</sup>, A M Muqsith<sup>5</sup>, M Alamsyah<sup>5</sup>

<sup>1</sup>School of Information System, Bina Nusantara University, Indonesia

<sup>2</sup>School of Computer Science, Bina Nusantara University, Indonesia

<sup>3</sup>Faculty of Humanities, Bina Nusantara University, Indonesia

<sup>4</sup>Research Interest Group in Education Technology, Bina Nusantara University, Indonesia

<sup>5</sup>School of Computer Science, Bina Nusantara University, Indonesia

**Abstract.** Government plays an important role in nation economic growth. Nevertheless, there are still many occurrences of government officers abusing their offices to do an act of corruption. In this order, the central government should pay attention to every area in the nation to avoid corruption case. Meanwhile, the news media always constantly preach about corruption case, this makes the news media relevant for being one of the sources of measurement of corruption perception index (CPI). It is required to map the corruption case in Indonesia so the central government can pay attention to every region in Indonesia. To develop the mapping system, researchers use Naïve Bayes Classifier to classify which news articles talk about corruption and which news articles are not, before implementing a Naïve Bayes Classifier there are some text processing such as tokenizing, stopping, and stemming.

## 1. Introduction

Government plays an important role in nation economic growth. Nevertheless, there are still many occurrences of government officers abusing their offices to do an act of corruption. Corruption becomes an undeniable disease and threatens the structure of society (Rahimi & Shakeri, 2016). Generally, corruption limits the governments' institution by disobeying procedure, use of resources, and office occupation without the legal procedure. Corruption also undermines the legitimacy of government and other democratic values (Ghosh & Siddique, 2014). As a national issue, corruption must be dealt with firm and clear with all the potential involvement that exists within the public especially the government and law enforcement officials. The concept of systemic corruption as a collective action has significant implications on policy implementation (Persson, Rothstein, & Teorell, 2013). The media can show the current conditions of the region.

Additionally, the media can also summarize thoughts and ideas of the public (Malik, 2001). Therefore, using the social media as source data to calculate Corruption Perception Index (CPI) is very possible. CPI measures the level of corruption and CPI mapping will help the government to analyze and take measures to prevent corruption in every region (Melgar, Rossi, & Smith, 2010; Rohwer, 2009). CPI may be a reference to map corruption in all regions in Indonesia; CPI mapping will help the government in the analysis and prevention of corruption in each region (Uca, Ince, & Sumen, 2016). Therefore, the researchers propose system development of corruption cases mapping based on Indonesia's Corruption Perception Index. To develop the mapping system, the researchers need to solve



the problem of how to classify news articles that talk about corruption case and news articles that do not talk about corruption case.

According to previous research, several algorithms can be used to classify the document including Naïve Bayes Classifier, Decision Trees, and Support Vector Machine, which is Naïve Bayes Classifier is the most accurate algorithm to classify the document (Huang, Lu, & Ling, 2003).

## 2. News Classification

### 2.1. Data Collection

The data collected in this research is a news content taken from some news site in Indonesia, there are some considerations made in the selection of news sites for data collection, the consideration include:

#### 2.1.1. Alexa rank.

Alexa rank is a website that calculates the rank of a website based on traffic by counting unique visitors with page views. Alexa rank produces good indicator to determine the credibility of a website (Olteanu, Peshterliev, Liu, & Aberer, 2013)

#### 2.1.2. News content that consistently uses 5W1H format

5W1H is an abbreviation of who, what, where, when, how and why. The information provided by an article will be very informative and clear if it can answer 5W1H (Jang & Woo, 2005)

#### 2.1.3. Site has existed since 2010

To equalize the final result of the calculation system, the site at least has published articles before or at least since 2010.

#### 2.1.4. Have sitemap feature that will be used for web crawling

Sitemap is a single page that displays a collection of structured links that directly connects the pages in a website, sitemap. With the sitemap, web crawling will easily index all the pages in a website (Yang et al., 2009)

#### 2.1.5. Site content can be retrieved using a scraping technique.

There are several websites that prohibit a web crawler to do indexing on the website so that the data of news item could not be retrieved.

There are seven Indonesian news sites considered for data collection. The news sites are Tempo (tempo.co), Detik (detik.com), Kompas (kompas.com), Merdeka (merdeka.com), Liputan6 (liputan6.com), Tribun News (tribunnews.com), and Sindo News (sindonews.com).

There are three techniques for data collection:

- Web Crawling is a software that runs recursively crawl from one page to another page by following the link in each page (Schrenk, 2007), in this research, crawling result will be saved in the database as a news article index for web scraping purpose.
- Web Scraping is a software technique that gathering and extracting information from a web page (Schrenk, 2007), in this research, web scraping is used to get news content based on news article index and saved the content into the database.
- Cron Scheduling is a UNIX and Linux base scheduling to run a command or a shell script periodically, in this research, cron scheduling is used to run web crawling and web scraping periodically.

By using web crawling, web scraping, and cron scheduling, 396565 news articles successfully collected, here is the detail of data collected:

*Table 1. Data Collected.*

News Site	Article Collected
Detik	66826
Kompas	49396
Liputan 6	87950

Merdeka	54475
Sindo News	41016
Tempo	60763
Tribun News	36139
<b>Total</b>	<b>396565</b>

## 2.2. News Classification Algorithm

There are two processes performed in the process of classifying the news content; the processes are text processing and news content classification process.

### 2.2.1. Text Processing

Text processing is done to represent the text into the same form (Tated & Ghonge, 2015). Text processing has three sub-processes; tokenizing, stopping and stemming:

- Tokenizing is a process that splits the document into a single word; this process split the document by space character and saved into the array of word.
- Stopping is a searching process that searches a stop word in the array of word and removes the found word from the array of word, an Indonesian stop word list is listed from Fadillah Z Tala’s research in 2003 (Tala, 2003).
- Stemming is a process that convert each word in the array of word into a root word for example “mencintai” into “cinta” or “memilikinya” into “milik”, the process used Nazief and Andriani’s algorithm, the algorithm is:

**Step 1**, search the word in a root word database, if found then the word is a root word and algorithm is finished.

**Step 2**, Inflection suffixes (-lah, -kah, -ku, -mu, or -nya) deletion, if a word is a particle (-lah, -kah, -tah, or -pun), then repeat this step to delete a possessive pronouns (-ku, -mu, or -nya) if exists.

**Step 3**, Derivation suffixes (-i, -an, or -kan) deletion, after deletion occurred, search the result word in a root word database, if found, then the algorithm is finished, if not then continue to Step 3.1.

**Step 3.1**, if -an was deleted and the last character was -k, then remove the -k, if the result found in root word database, then the algorithm is finished, if not then continue to Step 3.2.

**Step 3.2**, restore deleted suffixes (-i, -an, or -kan), and continue into Step 4.

**Step 4**, Derivation prefixes (di-, ke-, se-, me-, be-, pe-, or te-) deletion, with the maximum number of iterations is three times:

**Step 4.1**, the algorithm finished if:

- Not allowed prefixes and suffixes combination occurred based on this table (Table 2):

*Table 2. Not allowed prefixes and suffixes combination.*

Prefixes	Not allowed suffixes
be-	-i
di-	-an
ke-	-i, -kan
me-	-an
se-	-i, -kan

te- -an

- Current detected prefixes equal to last deletion prefixes.
- Three prefixes deletion.

**Step 4.2**, identify the prefixes type and delete, there are two types of prefixes:

- Standard prefixes (di-, ke-, or se-)
- Complex prefixes (me-, be-, pe-, or te-), this type can create a new word according to the followed root word, so, there are rule that is used to get the right hyphenation (Table 3):

**Table 3.** Hyphenation rule.

No	Word Format	Hyphenation
1	berV...	ber-V...   be-rV...
2	berCAP...	ber-CAP...   where C!=r and P!=r
3	berCAerV...	ber-CaerV... where C!=r
4	belajar	bel-ajar
5	beC1erC2...	be-C1erC2... where C1 != {rll}
6	terV...	ter-V...   te-rV...
7	terCerV...	ter-CerV ... where C!=r
8	terCP...	ter-CP... where C!=r
9	teC1erC2...	te-C1erC2... where C1!=r
10	me{llrlwly}V...	me-{llrlwly} V...
11	mem{blflv}...	mem-{blflv}...
12	mempe{rll}...	mem-pe...
13	mem{rV V}...	me-m{rV V}...   me-p{rV V}...
14	men{cld j z}...	men-{cld j z}
15	menV...	me-nV...   me-fV
16	meng{g l h q}...	meng-{g l h q}...
17	mengV...	meng-V...   meng-kV...
18	menyV...	meny-sV...
19	mempV...	mem-pV... where V!=e
20	pe{wly}V...	pe-{wly}V...
21	perV...	per-V...   pe-rV...
22	perCAP...	per-CAP... where C!=r dan P!=er
23	perCAerV...	per-CAerV... where C!=r
24	pem{blflv}...	pem-{blflv}...
25	pem{rV V}...	pe-m{rV V}...   pe-p{rV V}
26	pen{cld j z}...	pen-{cld j z}...
27	penV...	pe-nV...   pe-tV...
28	peng{g l h q}...	peng-{g l h q}
29	pengV...	peng-V...   peng-kV...
30	penyV...	peny-sV...
31	pelV...	pe-lV... exsept 'pelajar'
32	peCerV...	per-erV... where C!={rlwly llmln}
33	peCP...s	per-CP... where C!={rlwly llmln} and P!=er

- Search deleted prefixes result into a root word database, if not found then repeat the Step 4, if found, then the algorithm is finished.

**Step 5**, if word not found in root word database after Step 4, then repeat the Step 4, in this step, the algorithm is finished and assumed the word is a root word.

## 2.2.2. News Content Classification

### 2.2.2.1. Document Training

Documents that are trained in this process is the result of text processing, the document is 30 documents of corruption case, and 30 documents of the non-corruption case, each training result will be saved into a database, which will then be used for Naïve Bayes Classifier algorithm as a reference to an existing document.

### 2.2.2.2. Document Classification Algorithm

In this classification step, the used algorithm is Naïve Bayes Classifier algorithm, that uses a previous document to calculate the future probability using equation (Murphy, 2006):

$$V_{MAP} = \underset{V_j \in V}{\operatorname{arg\,max}} \prod_{i=1}^n (P(x_i|V_j)P(V_j))$$

$data$  Word from text processing

$x_i$  Data  $i$  (1,2,3, ...,  $n$ )

$V_j$  Class  $j$  (1,2,3, ...,  $n$ )

$P(x_i|V_j)$  Probability  $x_i$  to class  $V_j$ , where  $P(x_i|V_j) = \frac{n_k+1}{n+|data|}$

$P(V_j)$  Probability  $V_j$ , where  $P(V_j) = \frac{|documents\ j|}{|documents|}$

For  $P(V_j)$  and  $P(x_i|V_j)$  calculated while data training is being processed, where:

$|documents\ j|$  total documents in each class  $j$

$|documents|$  total documents in all class

$n_k$  total frequency of occurrence of each of data

$n$  total frequency of data appearance of each class

$|data|$  total all data from all class

## 3. Result and Analysis

To produce good results of Naïve Bayes classifier algorithm, there are 30 news articles that talk about corruption case and 30 news articles that do not talk about corruption case, then do the test of 10 articles that talk about corruption case and 10 articles that do not talk about corruption case, the result of the algorithm is 10 articles that talk about corruption was accurate 100%, and 10 articles that do not talk about corruption is accurate 100%.

Can be concluded, the research method that already described above can produce the accurate results. At this stage, the research can classify news articles that talk about corruption case or talk about non-corruption case.

**KPK Beri Isyarat Ratu Atut Terseret Kasus Korupsi**

Jakarta - Wakil Ketua Komisi Pemberantasan Korupsi (KPK) Busyro Muqoddas memberi sinyal terseretnya Gubernur Banten Ratu Atut Chosiyah dalam kasus dugaan korupsi. Meskipun tak menyebut secara gamblang kasus yang dimaksud, tapi menurut Busyro, Atut bisa jadi merupakan kepala daerah yang bisa diminta pertanggungjawaban.

"Ya, benar begitu, seperti Tangerang Selatan," kata Busyro di gedung kantornya, Senin, 18 November 2013. Sebelum bicara soal Atut, Busyro terlebih dahulu bicara soal adik ipar Atut yang juga Wali Kota Tangerang Selatan dalam kasus dugaan korupsi proyek pengadaan alat kesehatan di Pemerintah Kota Tangsel. (Baca: [Pelapar Dugaan Korupsi Atut Pemah Mau Dibuntut](#))

Menurut Busyro, saat ini dalam kasus alkes Tangsel, penyelenggara negara yang ditetapkan sebagai tersangka baru pada tingkat pejabat pembuat komitmen. "Cara kerja KPK, semua dimulai dari bawah, minggir-minggir-minggir, langsung nabrak ke atas," kata Busyro.

Busyro memberi contoh, dalam kasus dugaan korupsi PON Riau, Gubernur Riau Rusli Zainal bukan orang yang pertama ditetapkan menjadi tersangka. Dalam kasus travel cheque, Deputy Gubernur Senior Bank Indonesia Miranda Goeltom kena belakangan. "Ttu memang karakter kerja KPK. Tunggu saja, kami sedang mengumpulkan bukti," kata dia.

Terhitung 11 November 2013, KPK menetapkan tiga orang dalam kasus alkes Tangsel. Ketiganya adalah pejabat pembuat komitmen Mamak Jamaksari, peltinggi PT Mikindo Adiguna Pratama Dadang Prijatna, dan Chaerl Wardana alias Wawan, yang merupakan suami Alirin.

**MUHAMAD RIZKI**

---

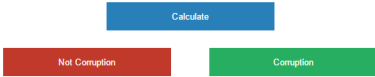


Figure 1. Training process of article that talks about corruption.

**Ribuan Truk Antre di Pelabuhan Merak**

Cilegon - Ribuan truk angkutan barang di Pelabuhan Merak. Kota Cilegon, Banten kembali ramai setelah sepuluh hari dilarang melintas pada H+4 sampai H+4 Lebaran. Meski tidak terjadi penumpukan, kendaraan truk tersebut mendominasi angkutan kapal dibandingkan jenis kendaraan lainnya seperti bus dan roda dua.

Manajer Operasional PT ASDP Indonesia Ferry Cabang Utama Merak, Zalis Anas mengatakan, angkutan truk diperkirakan masih akan terus terjadi hingga akhir pekan ini yaitu 10-11 September 2011. Ini terjadi lantaran banyak perusahaan yang sudah aktif kembali mengimpor barang baik dari arah Jawa ke Sumatera maupun sebaliknya. "Kami bersyukur arus mudik pada tahun ini berjalan lancar, lebih tertib dibandingkan tahun sebelumnya," kata Zalis, Kamis, 8 September.

Data dari PT Angkutan Sungai Danau dan Penyeberangan (ASDP) Indonesia Ferry Cabang Utama Merak menyebutkan, sejak H+5 truk mulai meningkat dari 302 unit menjadi 904 unit, pada H+6 jumlah truk melonjak menjadi 1.217 unit. Sedangkan jumlah bus menurun dari 532 pada H+5 menjadi 474 pada H+6, begitupun roda dua juga mengalami penurunan dari 1.411 pada H+5 menjadi 1.126 pada H+6. "Truk mulai ramai karena sebelumnya dilarang," kata Zalis Anas.

Kepala Dinas Perhubungan Komunikasi dan Informatika Provinsi Banten Husni Hasan mengatakan, penambahan demaga di Pelabuhan Merak, Banten dan Bakauheni, Lampung dari lima demaga menjadi enam demaga sangat mendesak dilakukan. Hal itu sebagai salah satu solusi untuk mengantisipasi penumpukan kendaraan akibat terus meningkatnya jumlah kendaraan yang melakukan penyeberangan.

"Kami sudah menyampalkannya kepada PT ASDP maupun ke Kementerian Perhubungan saat memantau di Merak. Solusi mengatasi kemacetan atau penumpukan kendaraan di Merak harus menambah demaga," kata Husni.

Menurutnya, berapa pun jumlah kapal yang disediakan PT ASDP Indonesia Ferry tidak akan mampu mengatasi penumpukan kendaraan. Karena jika dengan lima demaga itu, kapal yang dipasirkan tetap tidak bisa dimaksimalkan. "Penambahan jumlah kendaraan setiap tahunnya terus bertambah, apa lagi saat arus balik melebihi jumlah yang diprediksikan. Oleh sebab itu, perlu dilakukan langkah-langkah antisipasi dari sekarang dengan penambahan demaga," katanya.

Pantauan Tempo suasana arus balik di Pelabuhan Merak baik dari arah Sumatera maupun dari arah Jawa masih ramai. Loket penjualan tiket beberapa kali mengalami antrean, meski pun tidak sepanjang beberapa hari lalu. Namun, data dari PT ASDP pada H+6 masih mengangkut 57.562 penumpang dari Bakauheni ke Merak. Sedangkan dari Merak ke Bakauheni 45.605 penumpang. "Biasanya arus balik akan lebih lama, paling tidak lebih dari H+7," kata Kepala Cabang PT ASDP Merak, La Mane.

**WAS'UL ULUM**

---




Figure 2. Training process of article that does not talk about corruption.

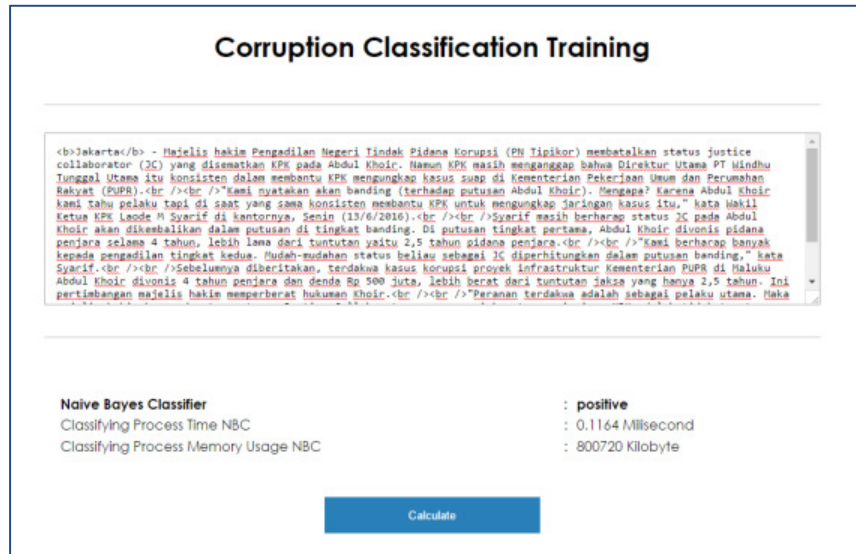


Figure 3. Testing process of article that talks about corruption.

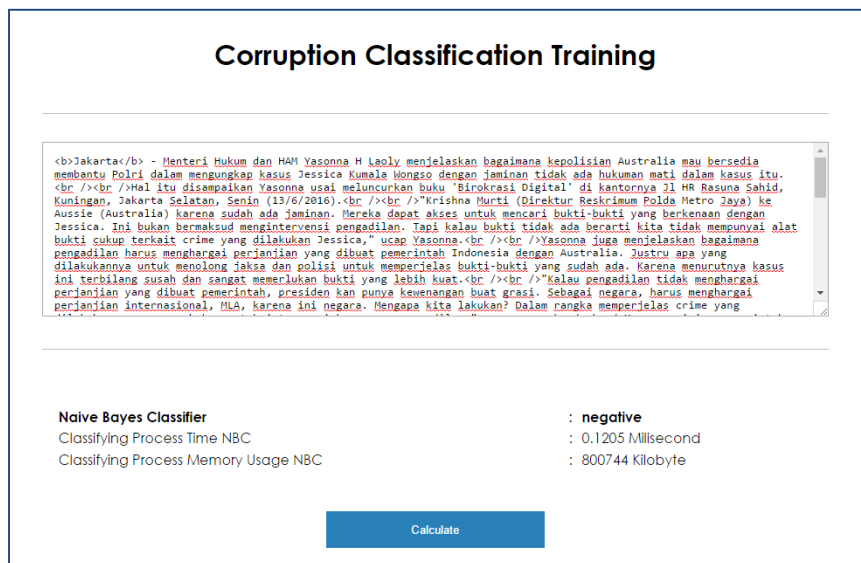


Figure 4. Testing process of article that does not talk about corruption.

From the process of text mining performed, resulting in the data on the mode of corruption based on the news phenomenon from online media related to corruption, which will be processed further to extract the location referred by the article. Thus, forming the data of the number of articles that have been classified by the scene, which can help the central government to perform certain actions on a particular place based on the data result of the text mining. It is expected to reduce cases of corruption in various regions in Indonesia.

## References

- [1] Ghosh, R., & Siddique, M. (2014). *Corruption, Good Governance and Economic Development: Contemporary Analysis and Case Studies*.

- [2] Huang, J., Lu, J., & Ling, C. (2003). *Comparing naive Bayes, decision trees, and SVM with AUC and accuracy*.
- [3] Jang, S., & Woo, W. (2005). 5W1H: Unified User-Centric Context. *Proceedings of the 7th International Conference on Ubiquitous Computing*.
- [4] Malik, D. (2001). *Dari Konstruksi ke Dekonstruksi: Refleksi atas Pemberitaan Televisi Kita*.
- [5] Melgar, N., Rossi, M., & Smith, T. W. (2010). The Perception of Corruption. *International Journal of Public Opinion Research*, 22(1), 120–131.
- [6] Murphy, K. P. (2006). *Naive Bayes Classifiers Generative Classifiers*, 4701, 1–8.
- [7] Olteanu, A., Peshterliev, S., Liu, X., & Aberer, K. (2013). Web Credibility: Features Exploration and Credibility Prediction. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 557–568.
- [8] Persson, A., Rothstein, B., & Teorell, J. (2013). Why Anti Corruption Reforms Fail — Systemic Corruption as a Collective Action Problem, *Governance*, 26(3), 449–471.
- [9] Rahimi, R., & Shakeri, H. (2016). Reflection on Judicial System's Corruption and Offering Solutions to Promote its Safety. *Journal of Politics and Law*, 9(9), 187.
- [10] Rohwer, A. (2009). Measuring Corruption: A Comparison between the Transparency International's Corruption Perception Index and the World Bank's Worldwide Governance Indicators. *DICE*, 42–52.
- [11] Schrenk, M. (2007). *Webbots, spiders, and screen scrapers*.
- [12] Tala, F. (2003). *A study of stemming effects on information retrieval in Bahasa Indonesia*
- [13] Tated, R. R., & Ghonge, M. M. (2015). A survey on text mining - techniques and application. *International Journal of Research in Advent Technology*, (1), 380–385.
- [14] Uca, N., Ince, H., & Sumen, H. (2016). The Mediator Effect of Logistics Performance Index on the Relation between Corruption Perception Index and Foreign Trade Volume. *European Scientific Journal*, 12(25), 1857–7881.
- [15] Yang, J. M., Cai, R., Wang, C., Huang, H., Zhang, L., & Ma, W. Y. (2009). Incorporating Site-Level Knowledge for Incremental Crawling of Web Forums: A List-Wise Strategy. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1375–1383.