

## International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 6.017

IJCSMC, Vol. 7, Issue. 8, August 2018, pg.159 – 169

# Indonesian Text News Classification Using the Naïve Bayes Algorithm

Ardiansyah Does<sup>1</sup>, Fajar Masya<sup>2</sup>, Hendra Prastiawan<sup>3</sup>

Department of Information System, Faculty of Computer Science, Universitas Mercu Buana

<sup>1</sup>[ardian@mercubuana.ac.id](mailto:ardian@mercubuana.ac.id), <sup>2</sup>[fmasya@mercubuana.ac.id](mailto:fmasya@mercubuana.ac.id), <sup>3</sup>[hendra.prastiawan@mercubuana.ac.id](mailto:hendra.prastiawan@mercubuana.ac.id)

*Abstract: Information has become a necessity in online social media life, from important things to ordinary things like gossip. The abundance of information that is on the Internet, makes someone enter the virtual world to find the information needed. The information needs are related to the selection of information keyword criteria to be appropriate and appropriate to the information seekers. This information can be in the form of Indonesian text news, but Indonesian text news on the internet is very abundant. So that, when the information seeker wants the news, it is done manually, which is like reading the title that has similarities, and the content of the news, then the information can be stored if it is in accordance with the information seeker's desire. But this method will use a lot of time to get the right information. This study classifies news from the official website of Okezone. The categories discussed are only Finance, Lifestyle, News and Sport. The document used for training data is 8 data which is then changed in the form of txt. The classification method used is naïve bayes classifier.*

*Keywords: News, Classification, Category, Naïve Bayes*

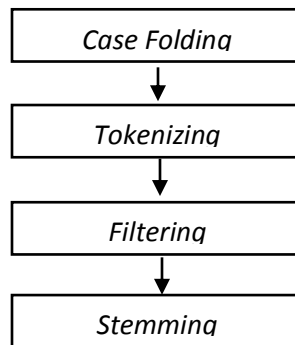
## INTRODUCTION

Information has become a necessity in online social media life, from important things to ordinary things like gossip. The information that develops on the internet is very abundant, so that when someone will search for information can enter the virtual world that there are thousands to millions of information needed.

A good data structure can facilitate the computerization process automatically. In text mining, information extracted contains information that is arbitrary in structure. Therefore, it is necessary to process the form into structured data according to their needs for processes in data mining, which will usually be numerical values. This process is often called Text Preprocessing, as is done by Ariadi and Fithriasari (Ariadi and Fithriasari, 2015) where they make changes to the data to facilitate processing. After the data becomes structured data and in the form of numerical values, the data can be used as a source of data that can be further processed.

Text that will be carried out by the text mining process, generally has several characteristics including having a high dimension, there is noise in the data, and there is a text structure that is not good. The method used in learning a text data, is to first determine the features that represent each word for each feature in the document. Before determining the features that represent, a preprocessing stage that is generally done in text mining in documents is

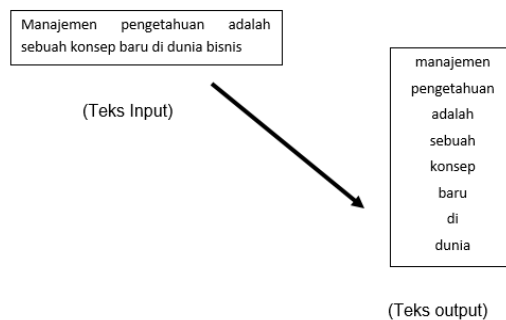
needed, namely case folding, tokenizing, filtering, and stemming. In Figure 1 the following shows document extraction in text preprocessing. (Efendi and Malik, 2012)



**Figure 1.** Dokumen extraction

Text in their raw form, however, is only a series of characters without explicit information about word and sentence boundaries. Before further processing can be done, the text needs to be segmented into words and sentences. This process is called tokenization. Tokenization divides the sequence of characters into sentences and sentences into tokens. Not only are words considered as evidence, but also numbers, punctuation marks, parentheses and quotation marks (Martiana et al., 2011).

According to Triawati (2009), the following picture 2 is an illustration of the case folding and tokenizing process.



**Figure 2.** Case Folding and Tokenizing sample process

The filtering stage is the stage of taking important words from the token results. Can use the stoplist algorithm (throw out less important words) or wordlist (save important words). Stoplist / stopword is non-descriptive words that can be discarded in the bag-of-words approach. Examples of stopwords are "the", "and", "on", "from" and so on (Martiana et al., 2011). Stemming stage is the stage of finding the root word of each word resulting from filtering. At this stage the process of returning various forms of words is carried out into the same representation. This stage is mostly used for English text and is more difficult to apply to Indonesian texts. This is because Indonesian does not have a permanent standard form formula. (Martiana et al., 2011).

The Naïve Bayes Classifier theory is part of the Bayes learning algorithm, where Bayes's decision theory is a fundamental statistical approach in pattern recognition (Santoso, 2007). While Wijaya and Santoso used the Naïve Bayes classification for e government sites (Wijaya and Santoso, 2016).

## RESEARCH METHODOLOGY

Classify Naïve Bayes here tries to find the highest probability value to classify test data in the most appropriate category. The stages in classify naïve bayes are:

1. For each word that appears from the test data, look into the word knowledge.
  - a. If there are, then look for probability values ( $P(w_k | v_j)$ ) for each category.

- b. If it's not there, then ignore the word.
- 2. For each category, calculate the value  $(P(v_j)) (P(w_k | v_j))$ .
- 3. After the multiplication results are obtained, the results of all categories will be compared and for the largest value, the news document will be included in that category. The Classify naïve Bayes process flow chart can be seen in Figure 5.2 below:

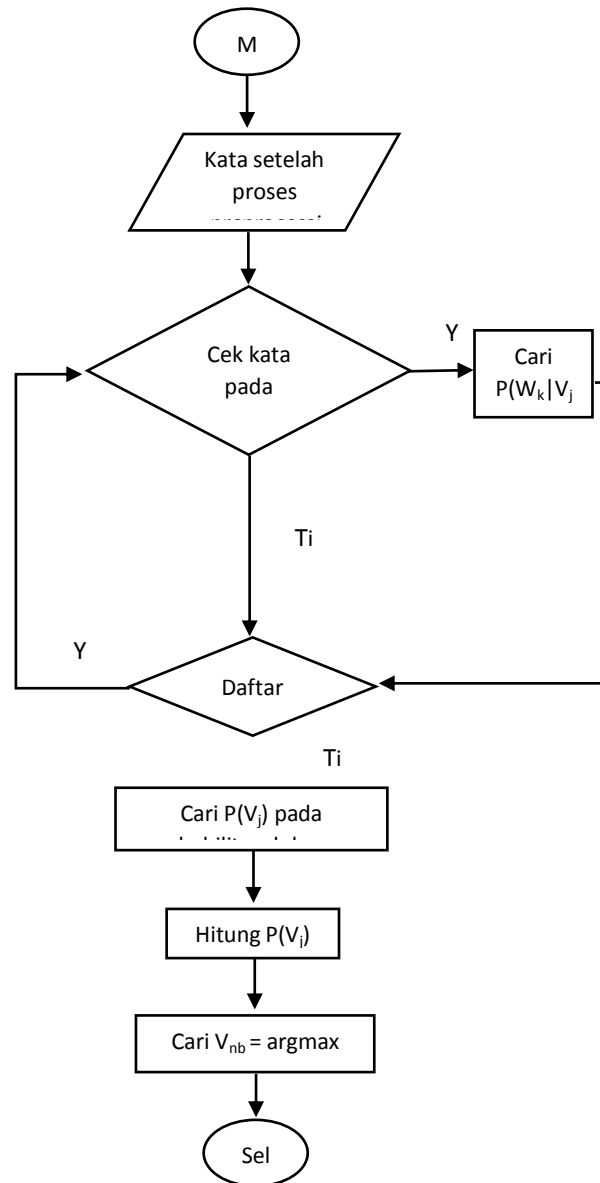


Figure 4.2. Classify Naive Bayes Process Flow Chart

## RESULTS AND DISCUSSION

### Data Processing Training

Training data that will be processed as many as 8 documents consisting of categories of Finance, Lifestyle, News, and Sport. Each category has 2 documents taken from the official site of Okezone. To get a probability list of word knowledge using the formula below:

$$\text{Formulation } P(w_k | v_j) = \frac{n_k + 1}{n + |\text{kosakata}|}$$

The following is a table listing knowledge of words that have been processed.

**Table 1. List of Knowledge Words**

No.	Kata	Frekuensi	P[Finance]	P[Lifestyle]	P[News]	P[Sport]
1	penerimaan	8	0.022058824	0.0026178	0.002604167	0.00253165
2	Pajak	5	0.014705882	0.0026178	0.002604167	0.00253165
3	Triliun	5	0.014705882	0.0026178	0.002604167	0.00253165
4	Target	4	0.012254902	0.0026178	0.002604167	0.00253165
5	anggaran	2	0.007352941	0.0026178	0.002604167	0.00253165
6	apbn	2	0.007352941	0.0026178	0.002604167	0.00253165
7	belanja	2	0.007352941	0.0026178	0.002604167	0.00253165
8	negara	2	0.007352941	0.0026178	0.002604167	0.00253165
9	pendapatan	2	0.007352941	0.0026178	0.002604167	0.00253165
10	realistis	2	0.007352941	0.0026178	0.002604167	0.00253165
11	analysis	1	0.004901961	0.0026178	0.002604167	0.00253165
12	center	1	0.004901961	0.0026178	0.002604167	0.00253165
13	Cita	1	0.004901961	0.0026178	0.002604167	0.00253165
14	cukai	1	0.004901961	0.0026178	0.002604167	0.00253165
15	capai	1	0.004901961	0.0026178	0.002604167	0.00253165
16	nilai	1	0.004901961	0.0026178	0.002604167	0.00253165
17	patok	1	0.004901961	0.0026178	0.002604167	0.00253165
18	direktur	2	0.007352941	0.0026178	0.002604167	0.00253165
19	eksekutif	1	0.004901961	0.0026178	0.002604167	0.00253165
20	indonesia	11	0.007228916	0.01028278	0.002604167	0.01772152
21	kepabeanaan	1	0.004819277	0.00257069	0.002604167	0.00253165
22	bus	4	0.012048193	0.00257069	0.002604167	0.00253165
23	stasiun	4	0.012048193	0.00257069	0.002604167	0.00253165
24	asian	4	0.009638554	0.00257069	0.002604167	0.00506329
25	games	4	0.009638554	0.00257069	0.002604167	0.00506329
26	gbk	3	0.009638554	0.00257069	0.002604167	0.01176471
27	gratis	3	0.009638554	0.00257069	0.002604167	0.01176471
28	kci	3	0.009638554	0.00257069	0.002604167	0.01176471
29	palmerah	3	0.009638554	0.00257069	0.002604167	0.01176471
30	bung	2	0.007228916	0.00257069	0.002604167	0.01176471
31	gelora	2	0.007228916	0.00257069	0.002604167	0.01176471
32	karno	2	0.007228916	0.00257069	0.002604167	0.01176471
33	transjakarta	2	0.007228916	0.00257069	0.002604167	0.01176471
34	kerjasama	1	0.004819277	0.00257069	0.002604167	0.01176471
35	operasi	1	0.004819277	0.00257069	0.002604167	0.01176471

36	commuter	1	0.004819277	0.00257069	0.002604167	0.01176471
37	dekat	1	0.004819277	0.00257069	0.002604167	0.01176471
38	disediakan	1	0.004819277	0.00257069	0.002604167	0.01176471
39	kawasan	1	0.004819277	0.00257069	0.002604167	0.01176471
40	kereta	1	0.004819277	0.00257069	0.002604167	0.01176471
41	layanan	1	0.004819277	0.00257069	0.002604167	0.01176471
42	lokasi	1	0.004819277	0.00257069	0.002604167	0.01176471
43	mengatakan	1	0.004819277	0.00257069	0.002604167	0.01176471
44	menit	1	0.004819277	0.00257069	0.002604167	0.01176471
45	menyediakan	1	0.004819277	0.00257069	0.002604167	0.01176471
46	rumah	5	0.002409639	0.00257069	0.002604167	0.01176471
47	pukul	3	0.002409639	0.00257069	0.002604167	0.01176471
48	masih	2	0.002409639	0.00257069	0.002604167	0.01176471
49	momen	2	0.002409639	0.00257069	0.002604167	0.01176471
50	orang	2	0.002409639	0.00257069	0.002604167	0.01176471
51	pagi	2	0.002409639	0.00257069	0.002604167	0.01176471
52	kedua	3	0.002409639	0.00257069	0.002604167	0.01176471
53	abadikan	1	0.002409639	0.00257069	0.002604167	0.01176471
54	alex	1	0.002409639	0.00257069	0.002604167	0.01176471
55	bersaudara	1	0.002409639	0.00257069	0.002604167	0.01176471
56	cikini	1	0.002409639	0.00257069	0.002604167	0.01176471
57	dini	1	0.002409639	0.00257069	0.002604167	0.01176471
58	frans	1	0.002409639	0.00257069	0.002604167	0.01176471
59	hari	1	0.002409639	0.00257069	0.002604167	0.01176471
60	jakarta	1	0.002409639	0.00257069	0.002604167	0.01176471
61	jalan	2	0.002409639	0.00514139	0.005208333	0.00253165
62	jam	1	0.002409639	0.00257069	0.002604167	0.00253165
63	jarum	1	0.002409639	0.00257069	0.002604167	0.00253165
64	jumat	1	0.002409639	0.00257069	0.002604167	0.00253165
65	kemerdekaan	1	0.002409639	0.00257069	0.002604167	0.00253165
66	lelap	1	0.002409639	0.00257069	0.002604167	0.00253165
67	masuk	2	0.002409639	0.00257069	0.002604167	0.00253165
68	mendur	1	0.002409639	0.00257069	0.002604167	0.00253165
69	mengenang	1	0.002409639	0.00257069	0.002604167	0.00253165
70	mengendap	1	0.002409639	0.00257069	0.002604167	0.00253165
71	menunjukkan	1	0.002409639	0.00257069	0.002604167	0.00253165
72	teh	6	0.002409639	0.01799486	0.002604167	0.00253165
73	hijau	4	0.002409639	0.01285347	0.002604167	0.00253165
74	tubuh	2	0.002409639	0.00771208	0.002604167	0.00253165

75	diminati	1	0.002409639	0.00514139	0.002604167	0.00253165
76	favorit	1	0.002409639	0.00514139	0.002604167	0.00253165
77	jenis	1	0.002409639	0.00514139	0.002604167	0.00253165
78	khasiatnya	1	0.002409639	0.00514139	0.002604167	0.00253165
79	membuatnya	1	0.002409639	0.00514139	0.002604167	0.00253165
80	minum	1	0.002409639	0.00514139	0.002604167	0.00253165
81	pencinta	1	0.002409639	0.00514139	0.002604167	0.00253165
82	terbaik	1	0.002409639	0.00514139	0.002604167	0.00253165
83	waktu	1	0.002409639	0.00514139	0.002604167	0.00253165
84	alami	3	0.002409639	0.00514139	0.002604167	0.00759494
85	antioksidan	1	0.002409639	0.00514139	0.002604167	0.00253165
86	bebas	1	0.002409639	0.00514139	0.002604167	0.00253165
87	berfungsi	1	0.002409639	0.00514139	0.002604167	0.00253165
88	bobot	1	0.002409639	0.00514139	0.002604167	0.00253165
89	dipercaya	1	0.002409639	0.00514139	0.002604167	0.00253165
90	kekebalan	1	0.002409639	0.00514139	0.002604167	0.00253165
91	kulit	1	0.002409639	0.00514139	0.002604167	0.00253165
92	mampu	1	0.002409639	0.00514139	0.002604167	0.00253165
93	presiden	3	0.002409639	0.01028278	0.002604167	0.00253165
94	anies	2	0.002409639	0.00771208	0.002604167	0.00253165
95	dewan	2	0.002409639	0.00771208	0.002604167	0.00253165
96	panjang	2	0.002409639	0.00771208	0.002604167	0.00253165
97	penetapan	2	0.002409639	0.00771208	0.002604167	0.00253165
98	prosesnya	2	0.002409639	0.00771208	0.002604167	0.00253165
99	sandiaga	2	0.002409639	0.00771208	0.002604167	0.00253165
100	surat	2	0.002409639	0.00771208	0.002604167	0.00253165
101	baik	1	0.002409639	0.00514139	0.002604167	0.00253165
102	baru	1	0.002409639	0.00257069	0.005208333	0.00253165
103	baswedan	1	0.002409639	0.00514139	0.002604167	0.00253165
104	dimulai	1	0.002409639	0.00514139	0.002604167	0.00253165
105	diproses	1	0.002409639	0.00514139	0.002604167	0.00253165
106	diri	1	0.002409639	0.00514139	0.002604167	0.00253165
107	mengundurkan	1	0.002409639	0.00514139	0.002604167	0.00253165
108	merinci	1	0.002409639	0.00514139	0.002604167	0.00253165
109	nulis	1	0.002409639	0.00514139	0.002604167	0.00253165
110	papar	1	0.002409639	0.00514139	0.002604167	0.00253165
111	paripurna	1	0.002409639	0.00514139	0.002604167	0.00253165
112	sukabumi	6	0.002409639	0.00257069	0.018229167	0.00253165
113	pelaku	3	0.002409639	0.00257069	0.010416667	0.00253165

114	alias	2	0.002409639	0.00257069	0.0078125	0.00253165
115	kelompok	2	0.002409639	0.00257069	0.0078125	0.00253165
116	kota	2	0.002409639	0.00257069	0.0078125	0.00253165
117	pembacokan	2	0.002409639	0.00257069	0.0078125	0.00253165
118	tangkap	2	0.002409639	0.00257069	0.0078125	0.00253165
119	video	2	0.002409639	0.00257069	0.0078125	0.00253165
120	akbp	1	0.002409639	0.00257069	0.005208333	0.00253165
121	aksi	1	0.002409639	0.00257069	0.005208333	0.00253165
122	antarkedua	1	0.002409639	0.00257069	0.005208333	0.00253165
123	barat	1	0.002409639	0.00257069	0.005208333	0.00253165
124	beredar	1	0.002409639	0.00257069	0.005208333	0.00253165
125	berhasil	1	0.002409639	0.00257069	0.005208333	0.00253165
126	cianjur	1	0.002409639	0.00257069	0.005208333	0.00253165
127	condro	1	0.002409639	0.00257069	0.005208333	0.00253165
128	ditangkap	1	0.002409639	0.00257069	0.005208333	0.00253165
129	gsi	1	0.002409639	0.00257069	0.005208333	0.00253165
130	jawa	1	0.002409639	0.00257069	0.005208333	0.00253165
131	kabupaten	1	0.002409639	0.00257069	0.005208333	0.00253165
132	kapolres	1	0.002409639	0.00257069	0.005208333	0.00253165
133	kasus	1	0.002409639	0.00257069	0.005208333	0.00253165
134	kata	1	0.002409639	0.00257069	0.005208333	0.00253165
135	media	1	0.002409639	0.00257069	0.005208333	0.00253165
136	memperlihatkan	1	0.002409639	0.00257069	0.005208333	0.00253165
137	ducati	6	0.002409639	0.00257069	0.002604167	0.01772152
138	pembalap	5	0.002409639	0.00257069	0.002604167	0.01518987
139	michelin	4	0.002409639	0.00257069	0.002604167	0.01265823
140	Ban	4	0.002409639	0.00257069	0.002604167	0.01265823
141	Bos	2	0.002409639	0.00257069	0.002604167	0.00759494
142	Dall	2	0.002409639	0.00257069	0.002604167	0.00759494
143	Igna	2	0.002409639	0.00257069	0.002604167	0.00759494
144	masalah	2	0.002409639	0.00257069	0.002604167	0.00759494
145	musim	2	0.002409639	0.00257069	0.002604167	0.00759494
146	peningkatan	2	0.002409639	0.00257069	0.002604167	0.00759494
147	performa	2	0.002409639	0.00257069	0.002604167	0.00759494
148	puas	2	0.002409639	0.00257069	0.002604167	0.00759494
149	manis	2	0.002409639	0.00257069	0.002604167	0.00759494
150	Tim	2	0.002409639	0.00257069	0.002604167	0.00759494
151	dilakukan	1	0.002409639	0.00257069	0.002604167	0.00506329
152	Gigi	1	0.002409639	0.00257069	0.002604167	0.00506329

153	kondisi	1	0.002409639	0.00257069	0.002604167	0.00506329
154	membuktikan	1	0.002409639	0.00257069	0.002604167	0.00506329
155	mengaku	1	0.002409639	0.00257069	0.002604167	0.00506329
156	menggunakan	1	0.002409639	0.00257069	0.002604167	0.00506329
157	menurutnya	1	0.002409639	0.00257069	0.002604167	0.00506329
158	grup	4	0.002409639	0.00257069	0.002604167	0.01265823
159	timnas	3	0.002409639	0.00257069	0.002604167	0.01012658
160	fase	2	0.002409639	0.00257069	0.002604167	0.00759494
161	laga	2	0.002409639	0.00257069	0.002604167	0.00759494
162	lolos	2	0.002409639	0.00257069	0.002604167	0.00759494
163	penggawa	2	0.002409639	0.00257069	0.002604167	0.00759494
164	perjuangan	2	0.002409639	0.00257069	0.002604167	0.00759494
165	angka	2	0.002409639	0.00257069	0.002604167	0.00759494
166	hong	2	0.002409639	0.00257069	0.002604167	0.00759494
167	klasemen	2	0.002409639	0.00257069	0.002604167	0.00759494
168	poin	2	0.002409639	0.00257069	0.002604167	0.00759494
169	garuda	1	0.002409639	0.00257069	0.002604167	0.00506329
170	gemilang	1	0.002409639	0.00257069	0.002604167	0.00506329
171	irfan	1	0.002409639	0.00257069	0.002604167	0.00506329
172	jaya	1	0.002409639	0.00257069	0.002604167	0.00506329
173	laos	1	0.002409639	0.00257069	0.002604167	0.00506329

### Data Testing Processing

Based on the knowledge of the words obtained by, then the following documents are tested:

Test Document:

*"Begini Cara Inalum Tingkatkan Produksi 1 Juta Ton Aluminium Holding Industri pertambangan, PT Indonesia Asahan Aluminium (Persero) atau Inalum terus bertransformasi untuk menjadi perusahaan kelas dunia (Fortune 500). "Kami memproduksi produk turunan aluminium berupa Billet dan Foundry Alloy, melakukan uji coba proyek optimalisasi dan up-grading tungku peleburan. Upaya lain yang dilakukan untuk menjadi perusahaan kelas dunia dengan finalisasi studi kelayakan untuk pengembangan smelter baru, ekspansi pelabuhan, pabrik Calcined Petroleum Coke (CPC) bekerjasama dengan Pertamina," kata Direktur Utama Inalum, Budi Gunadi Sadikin dalam keterangan tertulisnya di Jakarta,"*

**Table 2. Test document**

No.	Kata	Frekuensi
1	aluminium	3
2	inalum	3
3	dunia	2
4	kelas	2
5	perusahaan	2



6	alloy	1
7	asahan	1
8	baru	1
9	bekerjasama	1
10	bertransformasi	1
11	berupa	1
12	billet	1
13	budi	1
14	calcined	1
15	cara	1
16	coba	1
17	coke	1
18	cpc	1
19	dilakukan	1
20	direktur	1
21	ekspansi	1
22	finalisasi	1
23	fortune	1
24	foundry	1
25	grading	1
26	gunadi	1
27	holding	1

Then find the value of each category based on the many documents used as training data.

➤ Perhitungan  $P(v_j) = \frac{|docs_j|}{|Contoh|}$

➤ Diketahui :

Jumlah Dokumen Finance = 2

Jumlah Dokumen Lifesyle = 2

Jumlah Dokumen News = 2

Jumlah Dokumen Sport = 2

➤  $P(\text{Finance}) = 2/8 = 0,25$

➤  $P(\text{Lifestyle}) = 2/8 = 0,25$

➤  $P(\text{News}) = 2/8 = 0,25$

➤  $P(\text{Sport}) = 2/8 = 0,25$

Below is the probability value of each category based on table 2:

**Table 3. Glossary of the test document**

No.	Kata	Frekuensi	P[Finance]	P[Lifestyle]	P[News]	P[Sport]
1	aluminium	3	*	*	*	*`
2	inalum	3	*	*	*	*
3	dunia	2	*	*	*	*
4	kelas	2	*	*	*	*
5	perusahaan	2	*	*	*	*
6	alloy	1	*	*	*	*
7	asahan	1	*	*	*	*
8	baru	1	*	*	0.005208333	*
9	bekerjasama	1	*	*	*	*
10	bertransformasi	1	*	*	*	*
11	berupa	1	*	*	*	*
12	billet	1	*	*	*	*
13	budi	1	*	*	*	*
14	calcined	1	*	*	*	*
15	cara	1	*	*	*	*
16	coba	1	*	*	*	*
17	coke	1	*	*	*	*
18	cpc	1	*	*	*	*
19	dilakukan	1	*	*	*	0.005063291
20	direktur	1	0.007352941	*	*	*
21	ekspansi	1	*	*	*	*
22	finalisasi	1	*	*	*	*
23	fortune	1	*	*	*	*
24	foundry	1	*	*	*	*
25	grading	1	*	*	*	*
26	gunadi	1	*	*	*	*
27	holding	1	*	*	*	*

By using the formula below, a calculation is performed to find the probability of the word against the category.

- Probabilitas pada kategori Finance :

$$P(\text{direktur} | \text{Finance}) = 0,007352941$$

- Probabilitas pada kategori News:

$$P(\text{direktur} | \text{News}) = 0.005208333$$

- Probabilitas pada kategori Sport :

$P(\text{direktur} | \text{Sport}) = 0.005063291$   
Kemudian Cari Probabilitasnya =

$P(\text{Finance}) * P(\text{Dokumen Pengujian} | \text{Finance}) = 0,25 * 0,007352941 = 0.001838235$

$P(\text{News}) * P(\text{Dokumen Pengujian} | \text{News}) = 0,25 * 0,007352941 = 0.001302083$

$P(\text{Sport}) * P(\text{Dokumen Pengujian} | \text{Sport}) = 0,25 * 0,007352941 = 0.001265823$

Based on the probability value above, the test document is **categorized as Finance**

## CONCLUSIONS AND RECOMMENDATIONS

Based on the research conducted, the following conclusions can be drawn:

1. Research classifies legal online news, with categories of finance, lifestyle, news and sports.
2. The documents used as training were 8 pieces, each of which was divided into two categories.
3. Documents for testing data as much as one piece, which correctly classifies the news of the finance category.

## REFERENCES

- [1]. Ariadi, Dio., Fithriasari, Kartika., 2015, Klasifikasi Berita Indonesia Menggunakan Metode Naive BnaiveClassification dan Support Vector Machine dengan Confix Stripping Stemmer. Jurnal Sains dan Seni ITS, Volume 4 no 2.
- [2]. Efendi, Rusdi., Malik, Reza Firsandaya., Sari, Jeni Mila., 2012, Klasifikasi Dokumen Berbahasa Indonesia Menggunakan Naïve Bayes Classifie, Journal of Research in Computer Science and Application. Volume 1 no 1.
- [3]. Martiana E., Pribadi A.A., 2012, Pencarian Judul TA menggunakan Text Mining dan Metode Suffix Tree, Jurusan Teknik Informatika, ITS
- [4]. Santoso, Budi, 2007, Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis, Graha Ilmu, Yogyakarta.
- [5]. Triawati, Chandra 2009, Metode Pembobotan Statistical Concept Based untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia, Institut Teknologi Telkom Bandung.
- [6]. Wijaya, Akhmad Pandhu., Santoso, Heru Agus., 2016, Naïve Bayes Clasification pada klasifikasi Dokumen untuk Identifikasi Konten E- Government. Journal of Applied Intelligent System, Volume 1 no 1.