COMPARISON OF PORTER'S STEMMING ALGORITHM AND NAZIEF & ADRIANI'S STEMMING ALGORITHM IN DETERMINING INDONESIAN LANGUAGE LEARNING MODULES

Siti Tuhpatussania[1] ; Ema Utami[2] ; Anggit Dwi Hartanto[3]

Informatics Engineering Study Program
AMIKOM Yogyakarta University
Yogyakarta, Indonesia
www.amikom.ac.id
[1] tuhpatussania@students.amikom.ac.id; [2] ema.u@amikom.ac.id ; [3] anggit@amikom.ac.id

(*) Corresponding Author

Abstract — One of the methods used to improve the performance of text summarization to obtain complete information in a learning module is by transforming the words in a module into basic words or, in other words, through a steaming process. The steaming process in Indonesian language texts is more complicated/complex because there are word affixes that must be removed to get the root word (root word) of a word, so this research will compare the two stemming algorithms of Porter and stemming Nazief & Adriani in the learning module at Mataram University of Technology. The test results of the Nazief & Adriani stemming algorithm on an average process duration of 51.8 seconds with an average accuracy of 74.175%. In Porter's Algorithm, the average processing time is 16.875 seconds, with an accuracy of 73.225%.

Keywords: Indonesian stemming, porter algorithm, nazief & adriani algorithm.

Abstrak— *Salah satu cara yang digunakan untuk meningkatkan performa peringkasan teks agar mendapatkan informasi yang utuh pada sebuah modul pembelajaran yaitu dengan mentransformasi kata-kata dalam sebuah modul tersebut ke kata dasarnya atau dengan kata lain melalui proses stemming. Proses stemming pada teks berbahasa Indonesia lebih rumit/kompleks karena terdapat imbuhan kata yang harus dibuang untuk mendapatkan root word (kata dasar) dari sebuah kata sehingga pada penelitian ini akan membandingkan dua algoritma stemming Porter dan stemming Nazief & Adriani pada modul pembelajaran di Universitas Teknologi Mataram. Hasil pengujian dari algoritma stemming Nazief & Adriani pada durasi proses rata-rata 51,8 detik dengan akurasi rata-rata 74,175%. Sedangkan pada algoritma Porter durasi proses rata-rata 16,875 detik dengan akurasi 73,225%.*

Kata kunci : *stemming bahasa Indonesia, algoritma porter, algoritma nazief & adriani.*

INTRODUCTION

Source information for one of the students could be obtained in a module usually used as a medium of learning by students or female students. The learning module is the driving medium student for study in a manner independent as well as help the lecturer in conveying material to achieve the aim of learning(Laurensius Setyabudhi & Sanusi, 2019). However, in general, module learning contains hundreds of pages, which makes students reluctant to dig for further information, so the information you get is not accurately related to the courses taken, for the needed summary text from every module learning needs students to search for effective and efficient information. Effective means the user gets relevant information with entered *queries*. Efficient means time short search.

Stemming is one method used for upgrading performance summary text with the method transforming the words in a module learning to the basics for then the base word is given weight to achieve aim summary text that can represent the whole from the document original (Jatikusumo & Derajad Wijaya, 2021). The stemming algorithm will be different for every language (Simanjuntak, 2022). As examples of English and Indonesian have different morphology, the algorithm *stemming* needed will be different. The process of *stemming* the text is more Indonesian tricky/complex because a must-word affix is thrown away to get the root word (base word) from a word(Simanjuntak, 2022).

Word formation could be done with affixation, reduplication, and compositum(Harja Susetya & Harja Susetya, 2022). Affixation (e.g., "ajar" to be "belajar"), reduplication (e.g., "meja"

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 18, No. 2 September 2022 | DOI: 10.33480/pilar.v18i2.3940

becomes meja-meja), and composite (e.g., "mata" becomes "matahari"). Moreover, Indonesian has 35 official affixes in the dictionary big language Indonesia and affixes could form *prefixes* (prefixes), *suffixes* (suffixes), *confixes,* or absorbable *infixes* from the language java. With many Indonesian rules, the proper stemming algorithm was needed to define base words text Indonesian.

Study-related basic word determination or *stemming* Indonesian ever carried out by(Damayanti, 2022). The purpose of looking for a synonym is to reduce plagiarism. Search for the similarity of words is not a synonym but a similarity or a word close to the original word in comparison. In a study, this is the data used in two creation write scientific where document the will visit stages preprocessing up to stemming using algorithm Nazief & Adriani for getting the root word which is then the root word the used as reference level similarity two creation write scientifically. The stemming algorithm Nazief-Adriani produces more accuracy _ good because it could apply with an arrangement to return the word that has been stemmed against _ creation written scientifically.

A related study was carried out by (Mandar et al., 2020), who classified Indonesian news using Naïve Bayes with Porter stemming. The accuracy results of 15 news data use the second method, the produce mark percentage of 79% with Precision value, i.e., 1.000, Recall 0.6666, F-Measure 0.7951.

In previous research, the Porter stemming algorithm and Nazief & Adriani stemming had their respective advantages in determining basic words, but the stemming documents used were relatively small and neatly arranged, thus motivating this study to use the stemming algorithm in determining the base words of a learning module. The number of words in one document is an average of 15,000 and is not structured. Comparison of Porter's stemming algorithm and Nazief & Adriani's stemming algorithm in this study in order to be able to determine the algorithm that gets the best accuracy and duration of the steaming process in the Indonesian language learning module

MATERIALS AND METHODS

Stemming's research will use the method of descriptive-qualitative. The method to investigate inherent truth relative and theoretical and use hermeneutics as a Step for looking for meaning and interpretation(Zaluchu, 2021). At the same time, study descriptive is one method of examining the status of a group of humans, one subject, a set of conditions, a system of thinking, or even a class of current events (Aprilliwanto et al., 2021). Research is done with the qualitative method, usually served descriptively.

In research, this process involves data processing through stages of *preparation data* or preparing a dataset in the form of module learning with pdf and docx formats that are not locked. Next, the data preprocessing stages, namely the process of solving the word from document single be the root word, and the last Step evaluation, Process flow, and stages study could seen in figure 1.

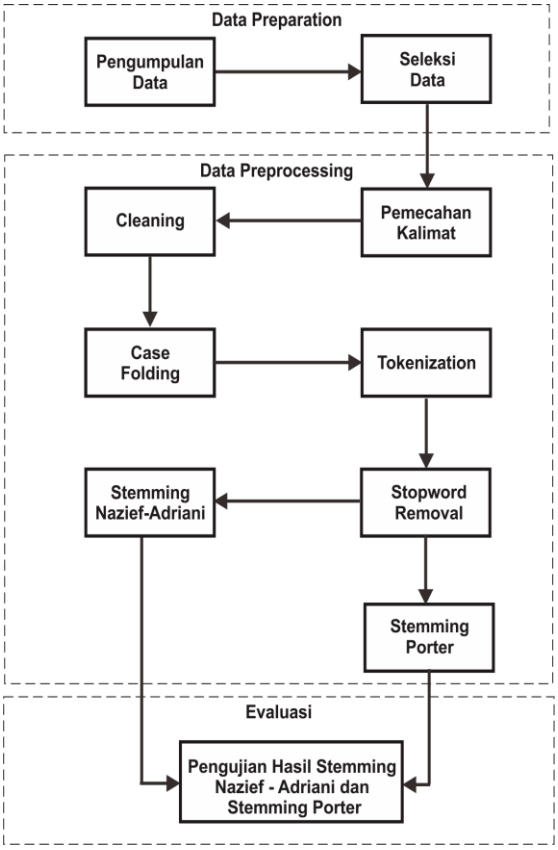

Figure 1. Stages Study

A. Data Preparation

Stages first is gathering module learning Indonesian language studies case of the University of Technology Mataram. The learning you get on the module originates from two faculty: Faculty Technology Information and Computers and Faculty Business and Law, which total 102 modules learning. The next module is selected based on the Indonesian language, and the files are not locked, so you get 20 modules learning from each faculty. Module list learning can be seen in table 1.

Table 1. List of Learning Modules

#	Title	Word Count	Faculty
1	Modul administrisi sistem.docx	21.059	FTIK
2	Modul KKPI.docx	4.357	FTIK
3	Modul komunikasi data.docx	14.021	FTIK
4	Modul pemrograman web I.docx	4.885	FTIK
5	Modul pemrograman web II.docx	5.883	FTIK
6	Modul praktikum IMK.docx	5.719	FTIK
7	Modul praktikum SBD.docx	3.518	FTIK
8	Modul sistem operasi.docx	8.831	FTIK
9	Modul technopreneurship 1.docx	8.952	FTIK
10	Modul rangkaian elektronika.pdf	7.725	FTIK
11	Modul mikroprosessor.pdf	13.451	FTIK
12	Modul algoritma dan pemrograman.pdf	29.673	FTIK
13	Modul aplikasi berbasis jaringan.pdf	16.224	FTIK
14	Modul instalasi komputer dan Troubelshooting.pdf	16.504	FTIK
15	Modul jaringan komputer lanjut.pdf	22.968	FTIK
16	Modul jaringan komputer.pdf	32.504	FTIK
17	Modul kecakapan antarpersonal.pdf	28.905	FTIK
18	Modul OOP.pdf	14.074	FTIK
19	Modul pemrograman mobile.pdf	27.614	FTIK
20	Modul technopreneurship ii.pdf	28.976	FTIK
21	Modul manajemen operasi perkantoran.docx	7.462	FBH
22	Modul efisiensi kantor.docx	6.959	FBH
23	Modul entrepreneurship 2.docx	8.184	FBH
24	Modul knb indonesia.docx	8.033	FBH
25	Modul komunikasi niaga.docx	6.688	FBH
26	Modul manajemen kearsipan.docx	12.463	FBH
27	Modul manajemen organisasi dan produksi.docx	9.874	FBH
28	Modul pengantar ilmu administrasi.docx	6.892	FBH
29	Modul SIM.docx	5.137	FBH
30	Modul statistik deskriptif.docx	7.851	FBH
31	Modul analisis laporan keuangan.pdf	17.751	FBH
32	Modul manajemen perkantoran.pdf	25.069	FBH
33	Modul metodologi penelitian bisnis.pdf	28.556	FBH
34	Modul penganggaran perusahaan.pdf	19.100	FBH
35	Modul statistik 1.pdf	16.898	FBH
36	Modul statistik 2.pdf	31.430	FBH
37	Modul studi kelayakan bisnis.pdf	29.203	FBH
38	Modul surat - menyurat.pdf	15.843	FBH
39	Modul UMKM dan koperasi.pdf	20.031	FBH
40	Modul-perpajakan.pdf	19.097	FBH

B. Preprocessing

After module learning has already been collected and selected, preprocessing is next. *Preprocessing* is converting raw data that is not structured and Becomes more structured (Rahman Isnain et al., 2021).

In research, there are several stages of preprocessing carried out that is as follows:

1. Solving sentence: this process is the Step beginning where text on the module cut learning _ be per sentence with sign separator used _ is (. , ? !) (Maulidia Sari & Siti Fatonah, 2021). Solving process sentences is also done in order data module Processed learning only from text with automatic delete that picture listed in the module learning.
2. *Cleaning*: in the stages of *cleaning*, the text is cleaned from *noises* like sign reads, numbers, and so on(Tuhpatussania et al., 2022). This also deletes URLs, hashtags, mentions, and omits character bytes because sign read the no can be processed internally stemming stage.
3. Case Folding: Case folding is the stage in changing all text. It Becomes alphabet small(Ridwansyah, 2022). Because the data on the module learning is not always structured and

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 18, No. 2 September 2022 | DOI: 10.33480/pilar.v18i2.3940

consistent, we needed stages for generalizing alphabet capital.

4. Tokenization: Tokenization in outline break sentence in something text to inrm a bunch of unit sentences for the counted result later (Joergensen Munthe et al., 2022). Stages this almost the same as stages solving sentence. The only stage this delimit or separator used is *space.*
5. Stemming: Stemming stages change the affix words becoming the word's basic form (Firman et al., 2022). in Indonesian, there are several commonly used stemming algorithms: Nazief & Adriani, Arifin & Setiono's Algorithm, Vega's Algorithm, Tala's Algorithm, and Porter's Algorithm.

Researchers use the algorithm Nazief & Adriani and Porter's Algorithm in Indonesian research for module learning. Where is the algorithm Nazief & Adriani Algorithm using a basic word dictionary (Saragih & Situmorang, 2019) meanwhile the Porter-based algorithm deletion affix (Rosidin et al., 2019)

Stemming Nazief & Adriani

Developed algorithm for the first time by Bobby Nazief and Mirna Adriani. This uses a basic word dictionary and does the recording, i.e., drafting back the words that experienced redundant stemming process(Merta Wirayasa et al., 2019). This Stemming Algorithm has stages as follows:

a. Stages look for the word to be stemmed in a basic dictionary if found so assumed word is the *root word* or base word. So algorithm stop.
b. *Inflexion Suffixes* ("- kah," "- lah," "-ku," "-nya," or "-mu" are removed. If in the form of *particles* ("-kah," "-lah," "-pun," or "-lah"), then step this repeated to remove pronouns ownership ("-ku," "- mu," or "-nya"), if there.
c. Wipe *Derivation Suffixes* ("- i," "-an" or "- kan"). The algorithm stops if the word is found in this stemming data dictionary. otherwise so to step c(1).

1) If the affix "-an" has been removed and the letters final from the word "k," then "-k" also follows deleted, and if the word there is in the dictionary, so the algorithm stops. Otherwise found, so carry on to step c(2).
2) Deleted suffixes ("-an," "-kan") are returned. Continue to step d.

d. Wipe *Derivation Prefix*. If in step point to c there is removed suffix _ so carry on to step d(1). If no, go to step d(2).

1) Check table combination. The prefix and the suffix are not allowed. Suppose found, so the algorithm stops. If not, carry on to Step d(2).
2) The equation "For i = 1 to 3" for determining the type later prefix _ prefix deleted. Suppose *the word root* is not yet found. Do Step on point e. If already, so the algorithm stops. Note: if the prefix second is the same as the prefix, the first algorithm is terminated.

e. Recoding.
f. If all Step has been done, but I cannot find the root word either, the source said the beginning would be considered the *root word*. Process complete. The advantage of the Algorithm Nazief & Adriani is that possible particles follow an affix word.

Stemming Porter

Porter's algorithm was first developed for stemming against the English language English, but this k has already been modified so it can be used on Indonesian stemming. Porter's Algorithm in Indonesian does disappearance particle in a word, for example, -lah in the word adalah, -kah in the word kamukah, removes the pronoun contained in a word, for example, -mu in the word bukumu, buku-nya, buku-ku, removes prefix one if no detected, then carry on to step-delete prefix second, and if there is so carry on to Step delete endings and stages end-remove prefix to two and continued to remove the suffix and delete suffix if permanent no found hence the word considered as basic words (Rosidin et al., 2019). Following description deletion addition to Porter's Algorithm.

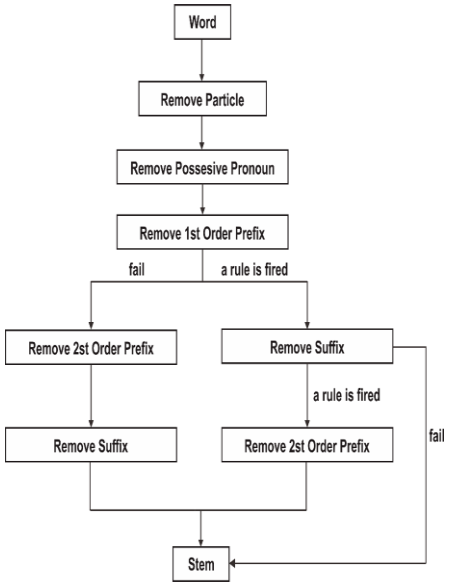

Figure 2. Porter's Algorithm

In Figure 2, there is a path for deleting affixes using Porter's stemming algorithm, where the particle removes the root word, the possessive

pronoun is deleted in the word, and the affix prefix is deleted. If there is no further prefix delete the second prefix, if the first prefix is found then continue to delete the suffix, if not found then the word is assumed to be the root word. If a suffix is found then remove the second prefix. Then the final word is assumed to be the root word

C. Evaluation

Nazief & Adriani's stemming results and Porter's stemming on modules learning next will do ratio accuracy after being tested using the Confusion Matrix. Because the Confusion Matrix is a table stating the classification of the number of correct test data and the number of wrong test data (Normawati & Prayogi, 2021) so could be applied to do testing for measure level Nazief & Adriani's stemming accuracy and Porter's stemming. Principles of Confusion Matrix in testing will count based on True Positive (TP), True Negative (TN), False Positive(FP), and False Negative (FP).

RESULTS AND DISCUSSION

After understanding the stages of determining the basic words in research, the next explanation results from the module stemming the shared learning process into two sub-chapters, like the following.

A. Process and Results of Preprocessing

Before Step preprocessing, the documents module learning from pdf and docx formats will be extracted as text, following example fragment paragraph on the Manajemen Operasi Perkantoran.docx module.

Table 2. Example of Module Raw Data

Fill
Menurut Quible (2002) yang dikutip oleh The Liang Gie(2007:161), layout menjelaskan penggunaan ruang secara efektif serta mampu memberikan kepuasan kepada pegawai terhadap pekerjaan yang dilakukan, maupun memberikan kesan yang mendalam bagi sipegawai

Table 3. Solving Results Sentences and Cleaning

Solving Sentence	Cleaning
“The Liang Gie(2007:161)”, “layout menjelaskan penggunaan ruang secara efektif serta mampu memberikan kepuasan kepada pegawai terhadap pekerjaan yang dilakukan”,”maupun memberikan kesan yang mendalam bagi sipegawai”	“The Liang Gie”, “layout menjelaskan penggunaan ruang secara efektif serta mampu memberikan kepuasan kepada pegawai terhadap pekerjaan yang dilakukan”,”maupun memberikan kesan yang mendalam bagi sipegawai”

Table 4. Result of Case Folding and Tokenization

Case Folding	tokenization
“the liang gie”, “layout menjelaskan penggunaan ruang secara efektif serta mampu memberikan kepuasan kepada pegawai terhadap pekerjaan yang dilakukan”,”maupun memberikan kesan yang mendalam bagi sipegawai”	“the”,”liang”,”gie”, “layout”,”menjelaskan”, ”penggunaan”,”ruang”, “secara”, “efektif”, “serta”, “mampu” “memberikan”, “kepuasan”, “kepada”, “pegawai”, “terhadap”, “pekerjaan”, “yang”, “dilakukan”, “maupun”, “memberikan”, “kesan” “yang”, “mendalam”, “bagi”, “sipegawai”

Table 5. Stemming Results

Nazief & Adriani	Porter
“the”,”liang”,”gie”, “layout”,”jelas”, “guna”, ”ruang”, “cara”, “efektif”, “serta”, “mampu” “beri”, “puas”, “pada”, “gawai”, “terhadap”, “kerja”, “yang”, “lakukan”, “maupun”, “beri”, “kesan” “yang”, “dalam”, “bagi”, “gawai”	“the”,”liang”,”gie”, “layout”,”jelas”, ”guna”,”ruang”, “secara”, “efektif”, “serta”, “mampu” “beri”, “puas”, “pada”, “gawai”, “hadap”, “kerja”, “yang”, “laku”, “maupun”, “beri”, “kesan” “yang”, “dalam”, “bagi”, “gawai”

Processing of basic words as shown in table 2 to table 4, namely the stages of raw data as in table 2 are paragraph fragments as an example and then split sentences based on separators (.) to get three sentences and then go through the cleaning stage, namely in the first sentence there are numbers and punctuation removed as in table 3. The next stage, namely in table 4, is Case Folding, wherein the first sentence contains capital letters because it is a person's name changed to all lowercase letters. Then the tokenization stage, which previously numbered three sentences, was split back into words which resulted in 26 words. The results of the separation of the words then go to the steaming process, as shown in Table 5. The base word will search for the word of origin by removing the affixes.

B. Testing

Stages end after the process of stemming is testing uses a *Confusion Matrix* to look for level accuracy stemming results as well measurements were also carried out duration (seconds) of the algorithm stemming process Nazief & Adriani and Porter's stemming algorithm with results as in table 6 below.

PILAR Nusa Mandiri: Journal of Computing and Information System
Vol. 18, No. 2 September 2022 | DOI: 10.33480/pilar.v18i2.3940

Table 6 Stemming Module using Algorithm Nazief & Adriani and Porter's Algorithm

Document	Nazief & Adriani		Porter	
	Duration	Accuracy	Duration	Accuracy
Module 1	47	84%	44	70%
Module 2	34	75%	12	76%
Module 3	20	77%	11	71%
Module 4	29	76%	11	76%
Module 5	42	75%	12	75%
Module 6	37	74%	12	64%
Module 7	25	73%	11	73%
Module 8	40	72%	11	72%
Module 9	29	71%	13	71%
Module 10	29	81%	11	78%
Module 11	49	73%	13	73%
Module 12	86	71%	22	71%
Module 13	83	75%	20	85%
Module 14	67	77%	27	77%
Module 15	90	70%	17	70%
Module 16	77	75%	22	75%
Module 17	87	57%	11	67%
Module 18	32	77%	14	77%
Module 19	99	82%	22	69%
Module 20	86	77%	33	77%
Module 21	35	76%	14	76%
Module 22	18	75%	10	75%
Module 23	27	74%	14	74%
Module 24	40	73%	11	73%
Module 25	35	72%	11	72%
Module 26	23	71%	11	71%
Module 27	43	75%	16	75%
Module 28	32	74%	13	76%
Module 29	32	73%	11	73%
Module 30	38	72%	16	72%
Module 31	34	71%	11	71%
Module 32	76	74%	25	74%
Module 33	89	72%	22	72%
Module 34	46	78%	21	71%
Module 35	52	75%	13	75%
Module 36	102	70%	30	73%
Module 37	73	71%	22	71%
Module 38	57	75%	12	75%
Module 39	100	82%	28	71%
Module 40	32	72%	15	72%

Table 6 shows the results accuracy and duration of the steaming process in each module learning with order according to the list in table 1. On the Algorithm Nazief & Adriani, the average processing time is 51.8 seconds with an average accuracy of 74.175%. In Porter's Algorithm, the average process duration is 16.875 seconds with 73.225% accuracy. Ratio stemming results can be seen in figure 3.

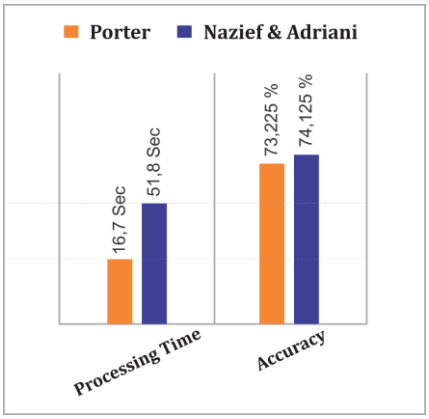

Figure 3 Percentage Comparison of Test Results Stemming Algorithm

CONCLUSION

Learning modules in Indonesian at the University of Technology Mataram have arranged complex sentences because arrangement can mean different. After all, sentences in the module are exposed from displayed images in a module that does not enter word *preprocessing*. However, results from the stemming algorithm can find *root wood* with ok. After testing using the Confusion Matrix, Porter's Stemming Algorithm and Nazief & Adriani's Stemming Algorithm each have superiority in determining the base word in the module learning Indonesian. In root word determination process time, Porter's stemming algorithm is faster, with a difference of 34.925 seconds from Nazief & Adriani's stemming algorithm. The average value of accuracy highest stemming results on the module learning the Indonesian language is 74.175% by Nazief & Adriani 's stemming algorithm or a difference of 0.95% from porter's algorithm.

REFERENCE

Aprilliwanto, R. E., Sanjaya, A., & Widodo, D. W. (2021). *Identifikasi Pola Kalimat Bahasa Indonesia Pada Siswa Sekolah Dasar Menggunakan Metode LALR dan Stemming*, *Seminar Nasional Inovasi Teknologi PGRI Kediri,* , e-ISSN: 2549-7952, p-ISSN: 2580-3336, 119-129. https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/915/615

Damayanti, A. (2022). Pendeteksi Untuk Tingkat Kemiripan Kata Pada Karya Tulis Ilmiah dengan Menggunakan Metode Nazief Adriani. *Bulletin of Information System Research (BIOS)*, 1(1), 30-38. https://journal.grahamitra.id/index.php/bios

Firman, S., Desena, W., Wibowo, A., Komputer, M. I., & Luhur, U. B. (2022). Penerapan Algoritma Stemming Nazief & Adriani Pada Proses Klasterisasi Berita Berdasarkan Tematik Pada Laman (Web) Direktorat Jenderal HAM Menggunakan Rapidminer. In *Syntax: Jurnal Informatika*, 11(2), 10-21. . https://doi.org/10.35706/syji.v11i02.7192

Harja Susetya, D. S., & Harja Susetya, H. H. (2022). Kesalahan Morfologi Bahasa Indonesia Pada Buletin Aktualita Lembaga Pers Mahasiswa Aspiratif Unzah. *Bahtera Indonesia; Jurnal Penelitian Bahasa Dan Sastra Indonesia*, *7*(2), 308–319. https://doi.org/10.31943/bi.v7i2.208

Jatikusumo, D., & Derajad Wijaya, H. (2021). Pendeteksi Lokasi Kejadian Covid-19 Menggunakan Social Media dengan Kombinasi Algoritma Stemming Bahasa Indonesia. In *IJCIT (Indonesian Journal on Computer and Information Technology)*, 6(1), 308-319. https://doi.org/https://doi.org/10.31294/ijcit.v6i1.10591

Joergensen Munthe, C. E., Astuti Hasibuan, N., & Hutabarat, H. (2022). RESOLUSI: Rekayasa Teknik Informatika dan Informasi Penerapan Algoritma Text Mining Dan TF-RF Dalam Menentukan Promo Produk Pada Marketplace. *Media Online*, *2*(3), 110–115. http://djournals.com/resolusi/article/view/309/223

Laurensius Setyabudhi, A., & Sanusi. (2019). Perancangan Modul Menggambar Teknik Berbasiskan Kompetensi Untuk Program Studi Teknik Industri Fakultas Teknik Ibnu Sina, . Jurnal Teknik Ibnu Sina, 4(2), 19-25. https://doi.org/10.3652/jt-ibsi.v4i2.41

Mandar, G., Muhamamd, A. H., Sudin, S., & Hamid, M. (2020). Klasifikasi Berita Indonesia Menggunakan Naïve Bayes dengan Porter Stemmer. *Jurnal Teknik Informatika (J-Tifa)*, *3*(2), 17–22. https://doi.org/10.52046/j-tifa.v3i2.1121

Maulidia Sari, Y., & Siti Fatonah, N. (2021). *Teks Otomatis pada Modul Pembelajaran Berbahasa Indonesia Menggunakan Metode Cross Latent Semantic Analysis (CLSA)*, *JEPIN (Jurnal Edukasi dan Penelitian Informatika) Peringkasan,* 7(2), 153-159. http://dx.doi.org/10.26418/jp.v7i2.47768

Merta Wirayasa, I. P., Agus Wirawan, I. M., & Ardi Pradnyana, I. M. (2019). *Algoritma Bastal: Adaptasi Algoritma Nazief & Adriani Untuk Stemming Teks Bahasa Bali*, Jurnal Nasional Pendidikan Teknik Informatika, 8(1), 60-69. https://doi.org/10.23887/janapati.v8i1.13500

Normawati, D., & Prayogi, S. A. (2021). Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter. In *Jurnal Sains Komputer & Informatika (J-SAKTI*), 5(2), 697-711.http://dx.doi.org/10.30645/j-sakti.v5i2.369

Rahman Isnain, A., Indra Sakti, A., Alita, D., & Satya Marga, N. (2021).), Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM, Jurnal Data Mining dan Sistem Informasi, 2(1), 31-37.

https://ejurnal.teknokrat.ac.id/index.php/JDMSI/article/view/1021

Ridwansyah, T. (2022). KLIK: Kajian Ilmiah Informatika dan Komputer Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier. Kajian Ilmiah Informatika dan Komputer, *2*(5), 178–185. http://djournals.com/klik/article/view/362/266

Rosidin, M., Fadlil, A., & Yudhana, A. (2019). Sistem Kelas Kata Berimbuhan menggunakan Algoritma Porter Stemmer Sebagai Pembelajaran Bahasa Indonesia, Telematika, 16(1), 11-17. https://doi.org/10.31315/telematika.v16i1.2988.g2439

Saragih, I. J., & Situmorang, Z. (2019). Perancangan Sistem Pelayanan Informasi Berbasis Android dengan Algoritma Nazief Andriani (StudiKasus :Universitas Katolik Santo Thomas), JUKI:Jurnal Komputer dan Informatika, 1(2), 66-77. https://ioinformatic.org/index.php/JUKI/article/view/19/16

Simanjuntak, I. Z. (2022). Analisa Kombinasi Algoritma Stemming Dan Algoritma Soundex Dalam Pencarian Kata Bahasa Indonesia. *Informasi Dan Teknologi Ilmiah (INTI)* , *10*(1), 24–30. http://ejurnal.stmik-budidarma.ac.id/index.php/inti/article/view/5040

Tuhpatussania, S., Utami, E., & Dwi Hartanto, A. (2022). Comparison Of Lexrank Algorithm And Maximum Marginal Relevance In Summary Of Indonesian News Text In Online News Portals, Jurnal Pilar Nusa Mandiri, 18(2), 187-192. https://doi.org/10.33480/pilar.v18i2.3190

Zaluchu, S. E. (2021). Metode Penelitian di dalam Manuskrip Jurnal Ilmiah Keagamaan. *Jurnal Teologi Berita Hidup*, *3*(2), 249-266. https://doi.org/10.38189/jtbh.v3i2.93