

**PERBANDINGAN AKURASI ALGORITMA *MULTINOMIAL NAÏVE BAYES (MNB)*
DAN *K-NEAREST NEIGHBOUR (KNN)* DALAM KLASIFIKASI ABSTRAK TUGAS
AKHIR
(STUDI KASUS: FAKULTAS TEKNIK UNIVERSITAS MUHAMMADIYAH JEMBER)**

Yuka Reksa Nugroho¹, Deni Arifianto, M.Kom.², Reni Umilasari, S.Pd, M.Si.³

Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember

Yukareksa1007@gmail.com¹, deniarifianto@unmuhjember.ac.id², reni.umilasari@gmail.com³

ABSTRAK

Tugas akhir merupakan salah satu syarat yang harus dipenuhi oleh mahasiswa untuk menyelesaikan pendidikan di suatu universitas. Pengelolaan kata yang terdapat pada abstrak memiliki suatu kemiripan antara abstrak satu dengan yang lainnya sehingga terjadi ambiguitas pengelompokan beberapa kategori. Hal tersebut juga membuat pengelompokan abstrak Tugas Akhir secara manual mengalami kesulitan. Tujuan dari penelitian ini adalah mengetahui tingkat akurasi seleksi atribut dalam klasifikasi abstrak tugas akhir. Pada penelitian ini dilakukan klasifikasi dokumen terhadap abstrak Tugas Akhir mahasiswa Fakultas Teknik Universitas Muhammadiyah Jember. Data yang digunakan dalam penelitian ini adalah abstrak Tugas Akhir pada Teknik Elektro, Teknik Sipil, Teknik Informatika, Teknik Mesin, dan Manajemen Informatika. Metode klasifikasi pada penelitian ini adalah membandingkan algoritma *Multinomial Naive Bayes* dan *K-Nearest Neighbour*, dikarenakan belum ada penelitian yang membandingkan algoritma MNB dan KNN dalam kasus klasifikasi abstrak tugas akhir.

Hasil pengujian dari penelitian ini secara keseluruhan, algoritma MNB di dapatkan hasil rata-rata akurasi sebesar 78%, rata-rata presisi sebesar 80%, dan rata-rata *recall* sebesar 80%. Sedangkan algoritma KNN mendapatkan hasil rata-rata akurasi sebesar 81%, rata-rata presisi sebesar 82%, dan rata-rata *recall* sebesar 82%.

Kata Kunci: Klasifikasi dokumen, Abstrak, *Multinomial Naive Bayes*, *K-Nearest Neighbour*

PENDAHULUAN

Latar Belakang.

Universitas Muhammadiyah Jember (UM Jember) merupakan salah satu perguruan tinggi swasta yang berada di Kota Jember. Universitas Muhammadiyah Jember memiliki sepuluh fakultas, salah satunya adalah Fakultas Teknik yang di dalamnya terdapat lima program studi yaitu Teknik Informatika, Teknik Sipil, Teknik Mesin, Teknik Elektro dan Sistem Informasi. Fakultas Teknik menerima mahasiswa baru setiap tahunnya. Banyaknya lulusan sarjana Fakultas Teknik yang akan dihasilkan tentunya berbanding lurus dengan banyaknya penelitian tugas akhir mahasiswa yang ada di Fakultas Teknik. Tugas akhir merupakan salah satu syarat yang harus dipenuhi oleh mahasiswa untuk menyelesaikan pendidikan di suatu universitas. Ilmu yang diperoleh selama perkuliahan akan diterapkan ke dalam suatu penelitian yang nantinya akan menghasilkan suatu keluaran berupa dokumen tugas akhir dan implementasi ilmu.

Seiring berkembangnya dunia informasi dan teknologi sering muncul permasalahan umum yaitu berkembang pesatnya jumlah data dan informasi yang ada. Perkembangan data dan informasi ini menyebabkan pengguna atau pencari informasi sedikit mengalami kesulitan karena terlalu besarnya data yang ada serta semakin lamanya waktu yang dibutuhkan untuk mencari sebuah data atau informasi dari pengorganisasian informasi atau dokumen yang diinginkan. Permasalahan ini juga terjadi pada klasifikasi dokumen pada data abstrak tugas akhir mahasiswa di

Fakultas Teknik Universitas Muhammadiyah Jember.

Berdasarkan uraian di atas, maka dibutuhkan sebuah metode klasifikasi dokumen / *Document Classification*. Permasalahan klasifikasi dokumen bisa diselesaikan dengan beragam metode, contohnya *Multinomial Naïve Bayes* dan *K-Nearest Neighbour*. *Multinomial Naïve Bayes* adalah model pembelajaran yang berorientasi dari Teorema Bayes. *Multinomial Naïve Bayes* sangat berguna untuk tugas-tugas pembelajaran yang melibatkan dimensi tinggi data, seperti klasifikasi teks & *web mining*. Sedangkan *K-Nearest Neighbour* adalah salah satu metode untuk melakukan klasifikasi berdasarkan *data training* atau data pembelajaran yang dilihat dari jarak yang paling dekat dengan objek berdasarkan banyaknya tetangga terdekat atau disebut juga dengan nilai K . Metode ini bertujuan mengklasifikasikan objek baru berdasarkan atribut dan *training sample*. Diberikan suatu titik *query*, selanjutnya akan ditemukan sejumlah K objek atau titik *training* yang paling dekat dengan titik *query*. Nilai prediksi dari *query* akan ditentukan berdasarkan klasifikasi tetangga (Tri, 2010).

Berdasarkan penelitian sebelumnya, yaitu *Online News Classification Using Multinomial Naïve Bayes* oleh Rahman dkk (2017), membahas tentang bagaimana proses klasifikasi pada berita *online*. Didapatkan hasil perhitungan dari Algoritma *Multinomial Naïve Bayes* dengan *TF-IDF* menghasilkan akurasi rata-rata tertinggi hingga 86,62% sedangkan *Multinomial Naïve Bayes* yang tidak menggunakan *TF-*

IDF Atau *DF-Thresholding* mencapai 86,28%, *Multinomial Naive Bayes* dengan *DF-Thresholding-TFIDF* hingga 86,15% dan *Multinomial Naive Bayes* dengan *DF-Thresholding* hingga 85,98%. Pada penelitian lainnya yaitu klasifikasi dokumen menggunakan Algoritma *K-Nearest Neighbour* oleh Nugraha (2018) membahas tentang klasifikasi dokumen untuk mendapatkan proses pencarian sebuah data atau dokumen yang valid dan sesuai dengan yang dibutuhkan. Berdasarkan hasil pengujian, algoritma KNN memiliki rata-rata akurasi sebesar 93,94438% pada klasifikasi dokumen. Perbandingan hasil klasifikasi dari algoritma *Naive Bayes* dan *K-Nearest Neighbour* pada penelitian Yusra (2016) yang membahas klasifikasi tugas akhir di jurusan Teknik informatika UIN Suska Riau didapatkan hasil bahwa algoritma *Naive Bayes* dapat menghasilkan akurasi yang lebih baik yaitu sebesar 87% sedangkan algoritma *K-Nearest Neighbour* menghasilkan menghasilkan nilai akurasi 84% dengan nilai $K = 3$, 85% dengan nilai $K = 5$, 86% dengan nilai $K = 7$ dan 84% dengan nilai $k=9$.

Berdasarkan uraian di atas, Algoritma *Multinomial Naive Bayes* dan *K-Nearest Neighbour* memiliki akurasi yang cukup baik dalam klasifikasi dokumen maka dari itu dalam penelitian ini akan dilakukan perbandingan algoritma klasifikasi pada data abstrak tugas akhir yang ada di fakultas Teknik Universitas Muhammadiyah Jember. Perbandingan terhadap kedua algoritma ini dilakukan agar dapat diketahui algoritma yang memiliki akurasi yang lebih baik terhadap pengklasifikasian abstrak tugas akhir. Pembaca sebuah tugas akhir

umumnya akan melihat kepada judul, abstrak dan daftar isi ketika membaca laporan tugas akhir. Dalam penelitian ini dilibatkan abstrak Bahasa Indonesia. Abstrak dipilih karena pada tugas akhir mahasiswa, abstrak merupakan rincian singkat dan jelas dari sebuah tugas akhir. Dari uraian tersebut maka penulis memutuskan penelitian ini berjudul “PERBANDINGAN AKURASI ALGORITMA MULTINOMIAL NAIVE BAYES DAN K-NEAREST NEIGHBOUR DALAM KLASIFIKASI ABSTRAK TUGAS AKHIR MAHASISWA FAKULTAS TEKNIK UNIVERSITAS MUHAMMADIYAH JEMBER”.

Rumusan Masalah

Berdasarkan latar belakang di atas, maka permasalahan yang akan dibahas dalam penelitian abstrak tugas akhir ini adalah berapa tingkat akurasi dari perbandingan algoritma *Multinomial Naive Bayes* dan *K-Nearest Neighbour* dalam klasifikasi abstrak tugas akhir Fakultas Teknik Universitas Muhammadiyah Jember?

Batasan Masalah

Agar permasalahan tidak menyimpang dari tujuan penelitian, maka berikut beberapa batasan yang perlu diperhatikan, yaitu :

1. Dataset yang digunakan pada penelitian ini adalah 100 data abstrak tugas akhir yang diambil dari *Repository* Universitas Muhammadiyah Jember.

2. Abstrak yang digunakan adalah abstrak tugas akhir mahasiswa Fakultas Teknik..
3. Jumlah data abstrak pada masing-masing kelas program studi yaitu 20 abstrak.
4. Metode pembobotan yang digunakan untuk dokumen abstrak pada penelitian ini adalah *TF/IDF*.
5. *Tools* yang digunakan pada penelitian ini adalah *Jupyter Notebook*.
6. Bahasa pemrograman yang digunakan pada penelitian ini adalah Bahasa pemrograman *Phyton*.
7. Output penelitian dikategorikan menjadi 5 kelas, yaitu : Prodi Teknik Elektro, Prodi Teknik Sipil, Prodi Teknik Informatika, Prodi Teknik Mesin, dan Prodi Manajemen Informatika

Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah untuk membandingkan tingkat akurasi algoritma *Multinomial Naive Bayes* dan *K-Nearest Neighbour* dalam klasifikasi abstrak tugas akhir mahasiswa Fakultas Teknik Universitas Muhammadiyah Jember.

Manfaat Penelitian

Manfaat dari penelitian ini yang dilakukan sebagai berikut .:

1. Bagi penulis
Hasil dari penelitian ini diharapkan dapat memberi kesempatan untuk menambah wawasan dan pengetahuan dalam bidang klasifikasi dokumen serta hasil

penelitian dapat di jadikan sebagai syarat kelulusan penulis.

2. Bagi instansi
Diharapkan dapat memberikan kontribusi secara keilmuan berupa hasil perbandingan algoritma *Multinomial Naive Bayes* dan *K-Nearest Neighbour* dalam klasifikasi abstrak tugas akhir mahasiswa Fakultas Teknik Universitas Muhammadiyah Jember.
3. Bagi peneliti lain
Hasil penelitian ini dapat digunakan sebagai referensi untuk penelitian-penelitian berikutnya di bidang klasifikasi dokumen dengan metode *text mining*.

TINJAUAN PUSTAKA

Pre-processing Text

Pre-processing text merupakan suatu proses pengubahan data berbentuk teks yang belum terstruktur menjadi data yang terstruktur sesuai dengan kebutuhan. Menurut Triawati dkk. (2009), tahapan pada *pre-processing* antara lain sebagai berikut:

- *Case folding*, merupakan proses mengonversi keseluruhan tulisan dalam data menjadi huruf kecil.
- *Tokenizing*, merupakan proses pemotongan tulisan berupa huruf untuk dipisahkan menjadi kata per kata.
- *Filtering*, merupakan proses menghapus kata-kata yang kurang penting dari hasil pemotongan tulisan.
- *Stemming*, merupakan proses mencari kata dasar dari tiap kata pembentuknya.

Pembobotan TF-IDF

Algoritma TF-IDF merupakan suatu metode pembobotan dengan menggunakan penghitungan *term frequency* dan *inverse document frequency*. Metode ini dilakukan dengan mencari representasi nilai dari tiap-tiap kata atau term pada sekumpulan data yang akan dibentuk menjadi sebuah vektor. Penggunaan TF-IDF dilakukan dengan cara pemberian bobot hubungan suatu kata atau fitur (t) sebanyak m terhadap data (d) sebanyak n , serta w merupakan hasil pembobotan TF-IDF.

Term Frequency

Term frequency (TF) merupakan frekuensi suatu term atau kata yang terdapat pada setiap data berbentuk dokumen. Nilai TF diperoleh berdasarkan jumlah kemunculan term tersebut pada setiap dokumen.

Inverse Document Frequency

Inverse document frequency (IDF) menunjukkan hubungan ketersediaan suatu term dalam seluruh dokumen (Harviant dan Kreinovich, 2014). Rumus yang digunakan dalam *inverse document frequency* ditunjukkan pada persamaan 1 (Harviant dan Kreinovich, 2014).

$$idf(t, d) = \log \frac{N}{df(t)} \quad (1)$$

Keterangan:

$idf(t, d)$ = IDF pada term t di dalam data d .

N = jumlah data.

$df(t)$ = jumlah data yang mengandung term t .

Term Weighting TF-IDF

Term weighting TF-IDF merupakan penggabungan dari rumus *term frequency* dengan *inverse document frequency* dengan mengalikan kedua rumus tersebut menjadi sebuah nilai pembobotan. Rumus TF-IDF ditunjukkan pada persamaan 2 (Harviant dan Kreinovich, 2014).

$$w(t, d) = tf(t, d) \times idf(t, d) \quad (2)$$

Keterangan:

$w(t, d)$ = hasil pembobotan TF-IDF pada term t di dalam data d .

$tf(t, d)$ = nilai kemunculan term t di dalam data d .

Multinomial Naïve Bayes

Multinomial Naive Bayes adalah probabilitas bersyarat yang dilakukan tanpa memperhitungkan urutan kata dan informasi yang ada dalam kalimat atau dokumen secara umum. Algoritma ini juga memperhitungkan jumlah kata yang muncul dalam dokumen (Destuardi dan Surya, 2009).

Misal terdapat dokumen d dan himpunan kelas c . Untuk memperhitungkan kelas dari dokumen d , maka dapat dihitung dengan rumus :

$$P(c | \text{term dokumen } d) = P(c) \times P(t_1 | c) \times P(t_2 | c) \times P(t_3 | c) \times \dots \times P(t_n | c) \quad (3)$$

Keterangan :

$P(c)$: Probabilitas *prior* dari kategori c .

t_n : Kata dokumen d ke- n

$P(c | \text{term})$: Probabilitas suatu

dokumen d) dokumen termasuk kategori *c*
 $P(t_n|c)$: Probabilitas kata ke-*n* dengan diketahui kategori *c*

Sementara rumus Multinomial yang digunakan dengan pembobotan kata TF-IDF adalah sebagai berikut :

$$P(w_k|v_j) = \frac{W_{ct} + 1}{(\sum_{W' \in VW'_{ct}} W') + B'} \quad (4)$$

Keterangan :

$P(w_k|v_j)$: Probabilitas kemunculan kata *wk* pada suatu dokumen jika diketahui dokumen tersebut berkategori tertentu.

W_{ct} : Nilai pembobotan *tfidf* atau *W* dari term *t* di kategori *c*.

$(\sum_{W' \in VW'})$: Jumlah total *W* dari keseluruhan term yang berada di kategori *c*.

B' : Jumlah *W* kata unik (nilai *idf* tidak dikali dengan *tf*) pada seluruh dokumen.

K-Nearest Neighbor

Algoritma *k-nearest neighbor* (KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. KNN bekerja dengan mencari sejumlah *k*-objek data atau pola (dari semua pola latih yang ada) yang paling dekat dengan pola masukan, kemudian memilih kelas dengan jumlah pola terbanyak di antara *k* pola tersebut (Suyanto, 2018). Dekat atau jauhnya lokasi (jarak) bisa dihitung melalui salah satu dari besaran

jarak yang telah ditentukan yakni jarak *Euclidean*, jarak *Minkowski*. Namun dalam penerapannya seringkali digunakan jarak *Euclidean* karena memiliki tingkat akurasi dan juga *productivity* yang tinggi. Jarak *Euclidean* adalah besarnya jarak suatu garis lurus yang menghubungkan antar objek. Rumus jarak *Euclidean* adalah sebagai berikut:

$$d(x_i, x_j) = \sqrt{\sum_{n=1}^p (x_{ip} - x_{jp})^2} \quad (5)$$

Dari rumus di atas dijelaskan bahwa, $d(x_i, x_j)$ merupakan jarak *euclidean* dari data test dengan data training sedangkan x_{ip} dan x_{jp} merupakan data testing ke *i* dan data training ke *j*. Optimalnya nilai *k* pada algoritma ini tergantung pada sebuah data. Secara umum, nilai *k* yang tinggi akan mengurangi efek *noise* pada klasifikasi, tetapi membuat batasan antara setiap klasifikasi menjadi semakin kabur.

HASIL DAN PEMBAHASAN

Data

Data yang digunakan berasal dari abstrak Tugas Akhir Fakultas Teknik Universitas Muhammadiyah Jember yang ada *repository* Universitas Muhammadiyah Jember. Teknik pengumpulan data yang digunakan adalah studi literatur dengan cara mengambil abstrak Tugas Akhir mahasiswa Fakultas Teknik Universitas Muhammadiyah Jember. Dari hasil studi literatur didapatkan data abstrak Tugas Akhir mahasiswa Fakultas Teknik Universitas Muhammadiyah Jember berjumlah 100 data abstrak Tugas Akhir, dengan masing-masing kelas atau prodi memiliki data abstrak berjumlah 20 abstrak.

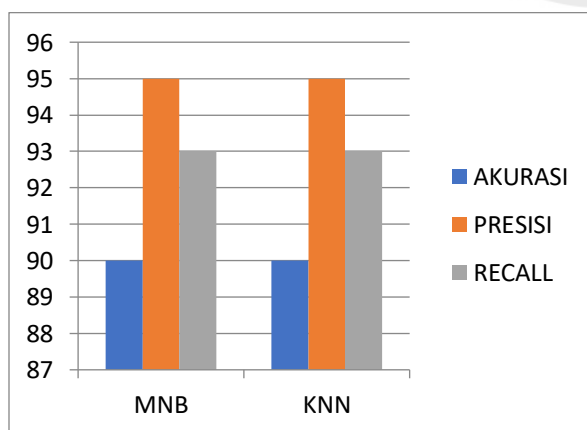
Pre-processing Data

Pre-processing data abstrak dilakukan teknik *case folding*, *tokenizing*, *filtering*, dan *stemming* pada setiap abstrak Tugas Akhir tersebut. Kemudian dilakukan pelabelan ulang pada data abstrak.

Hasil Klasifikasi

Hasil akurasi dengan *fold* $K = 2, 4, 5$ dan 10 pada *dataset* abstrak Tugas Akhir, untuk algoritma *Multinomial Naïve Bayes* didapatkan hasil akurasi tertinggi sebesar 90% pada nilai *fold* $K = 10$, skenario ke 1, presisi tertinggi sebesar 95% pada nilai *fold* $K = 10$, skenario ke 1 dan *recall* tertinggi sebesar 93% pada nilai *fold* $K = 10$, skenario ke 7, sedangkan untuk algoritma *K-Nearest Neighbour* didapatkan hasil akurasi tertinggi sebesar 90% pada nilai *fold* $K = 10$, skenario ke 10, presisi tertinggi sebesar 95% pada nilai *fold* $K = 10$, skenario ke 10 dan *recall* tertinggi sebesar 93% pada nilai *fold* $K = 10$, skenario ke 10.

Dari rata-rata secara keseluruhan, algoritma MNB di dapatakan hasil akurasi sebesar 78%, presisi sebesar 80%, dan *recall* sebesar 80%. Sedangkan algoritma KNN mendapatkan hasil akurasi sebesar 81%, presisi sebesar 82%, dan *recall* sebesar 82%.



KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat diambil kesimpulan sebagai berikut:

1. Hasil akurasi paling tinggi dalam klasifikasi Abstrak Tugas Akhir menggunakan algoritma *Multinomial Naïve Bayes* sebesar 90%, akurasi tersebut didapatkan pada *fold* $K=10$ skenario ke-1. Skenario ini juga menghasilkan nilai presisi sebesar 95% dan nilai *recall* sebesar 90%. Maka disimpulkan skenario ini adalah skenario terbaik dari algoritma *Multinomial Naïve Bayes*.
2. Hasil akurasi paling tinggi dalam klasifikasi abstrak Tugas Akhir menggunakan algoritma *K-Nearest Neighbour* sebesar 90%, akurasi tersebut didapatkan pada *fold* $K=10$ skenario ke-10. Skenario ini juga menghasilkan nilai presisi sebesar 95% dan nilai *recall* sebesar 93%. Maka disimpulkan skenario ini adalah skenario terbaik dari algoritma *K-Nearest Neighbour*.
3. Secara keseluruhan kinerja algoritma *K-Nearest Neighbour* lebih baik dibandingkan dengan *Multinomial Naive Bayes* dalam klasifikasi Abstrak Tugas Akhir, dapat disimpulkan seperti di atas dikarenakan hasil keseluruhan akurasi pada algoritma *K-Nearest Neighbour* lebih tinggi jika di rata-rata yaitu 81%.

Saran

Berdasarkan penelitian yang telah dilakukan, beberapa saran yang dapat dikembangkan untuk penelitian selanjutnya adalah sebagai berikut:

1. Penelitian perlu dikembangkan menggunakan data abstrak dengan jumlah yang lebih banyak, agar mendapatkan lebih banyak kata untuk diolah dan diklasifikasi serta agar dapat lebih akurat dalam pelabelan data sehingga hasil penelitian lebih maksimal.
2. Penelitian selanjutnya dapat membuat sebuah aplikasi berdasarkan hasil akhir dari penelitian ini. Aplikasi dapat dimanfaatkan untuk membantu mahasiswa di perpustakaan Fakultas Teknik Universitas Muhammadiyah Jember.

DAFTAR PUSTAKA

- Aisyah, S dan Fithriasari, K. (2016). "Klasifikasi Berita Online Menggunakan Metode Support Vektor Machine dan K-Nearest Neighbour". *JURNAL SAINS DAN SENI ITS*, Vol. 5 No. 2. Institut Teknologi Sepuluh Nopember.
- Destuardi dan Surya, S. 2009. "Klasifikasi Emosi Untuk Teks Bahasa Indonesia Menggunakan Metode Naive Bayes". Surabaya: Teknik Elektro, Institut Teknologi Sepuluh Nopember.
- Hall, M., 2006. A Decision Tree-Based Attribute Weighting Filter for Naive Bayes. *Knowledge-Based Systems*, pp.120–126. Available at: <http://www.cs.waikato.ac.nz/pubs/wip/2006/uow-cs-wp-2006-05.pdf>
- Harvlant, L. dan Kreinovich, V. 2014. *A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation)*. University of Texas at El Paso.
- Jo, T. 2018. *Text Mining: Concepts, Implementation, and Big Data Challenge*. Switzerland: Springer International Publishing AG.
- Kalokasari, D.H., dan Setyaningrum, A.H. 2017. Implementasi Algoritma Multinomial Naive Bayes Classifier Pada Sistem Klasifikasi Surat Keluar (Studi Kasus : DISKOMINFO Kabupaten Tangerang). *JURNAL TEKNIK INFORMATIKA* VOL.10 NO.2, 2017. UIN Syarif Hidayatullah, Jakarta.
- Kamus Besar Bahasa Indonesia (KBBI) [online]. Tersedia di <https://kbbi.web.id/abstrak>. Diakses 22 Nopember 2019.
- Miller, T. 2005. *Data and Text Mining: A Business Applications Approach*. Pearson Prentice Hall. ISBN 0131229117, 9780131229112.
- Mulyana, I., Ramadona, S., dan Herfina. 2012. "Penerapan Term Frequency-Inverse Document Frequency pada Sistem Peringkasan Teks Otomatis Dokumen Tunggal Berbahasa Indonesia".
- Nugraha, P.D., Faraby, S.A., dan Adiwijaya. 2018. "Klasifikasi Dokumen Menggunakan Metode K-Nearest

- Neighbour (KNN) dengan Information Gain”. *e-Proceeding of Engineering* : Vol.5, No.1. ISSN : 2355-9365. Prodi S1 Teknik Informatika, Fakultas Informatika, Universitas Telkom.
- Perpuspedia – Abstract : Universitas Diponegoro Digital Library [online]. Tersedia di <https://digilib.undip.ac.id/v2/2012/05/14/abstract/>. Diakses Januari 2020.
- Rahman, A. Wiranto dan Doewes, A. 2017. “Online News Classification Using Multinomial Naive Bayes”. ITSMART Vol. 6 No.1. Universitas Sebelas Maret.
- Repository Universitas Muhammadiyah Jember [online]. Tersedia di <http://repository.unmuhjember.ac.id/view/divisions/information/>. Diakses Nopember 2019.
- Setiawan, P. 2019. Pengertian Dan Manfaat Kalimat Abstrak Lengkap [online]. Tersedia di <https://www.gurupendidikan.co.id/kalimat-abstrak/>. Diakses Desember 2019.
- Suyanto. 2018. *Machine Learning: Tingkat Dasar dan Lanjutan*. Informatika Bandung, Bandung. ISBN 978-602-6232-78-6.
- Sya’bani. 2019. “Penerapan Metode Cosine Similarity dan Pembobotan TF/IDF pada Sistem Klasifikasi Sinopsis Buku di Perpustakaan Kejaksaan Negeri Jember”. JUSTINDO (Jurnal Sistem & Teknologi Informasi Indonesia). Universitas Muhammadiyah Jember.
- Taheri, S., Mammadov, M. and Bagirov, A. (2011). Improving naive Bayes classifier using conditional probabilities. 9th Australasian Data Mining Conference, Ballarat, Australia, pp. 63-68
- Triawati, dkk. 2009. “Metode Pembobotan Statistical Concept Based untuk Klastering dan Kategorisasi Dokumen Berbahasa Indonesia”. Institut Teknologi Telkom, Bandung.
- Tri, R., Prastiyanto, D., dan Suprpto, E. 2010. “Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi”. Jurnal Teknik Elektro Vol. 9 No. 1, P-ISSN 1411 – 0059, E-ISSN 2549 – 1571. Jurusan Teknik Elektro, Fakultas Teknik, Universitas Negeri Semarang.
- Weissberg, R and Buker, S. 1990. *Writing up Research: Experimental Research Report Writing for Students of English*. New Jersey: PrenticeHall, Inc. Dalam Yanti, Y. 2014. “Cara Penulisan Abstrak Di Dalam Karya Ilmiah”. Fakultas Ilmu Budaya (FIB) Universitas Bung Hatta.
- Yusra, Olivita, D., dan Vitriani, Y. 2016. “Perbandingan Klasifikasi Tugas Akhir Mahasiswa Jurusan Teknik Informatika Menggunakan Metode Naïve Bayes Classifier dan K-Nearest Neighbour”. *Jurnal Sains, Teknologi dan Industri*, Vol. 14, No. 1, pp. 79 – 85. ISSN 1693-2390. Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim Riau.

Zakky. 2018. Pengertian Abstrak Menurut Para Ahli, KBBI dan Secara Umum [online]. Tersedia di <https://www.zonareferensi.com/pengertian-abstrak/>. Diakses Januari 2020.

