

Perbandingan Metode *Naïve Bayes* dengan Metode *K-NN* untuk Klasifikasi Berita

*Diajukan Sebagai Syarat untuk Menyelesaikan
Pendidikan Program Strata-1 pada
Jurusan Teknik Informatika*



Oleh :

Muhamad Khaidiramsyah

09121402058

**Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2019**

LEMBAR PENGESAHAN TUGAS AKHIR

Perbandingan Metode *Naïve Bayes* dengan Metode *K-NN* untuk Klasifikasi Berita

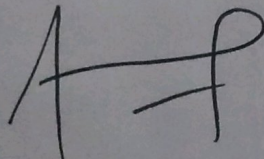
Oleh :

Muhamad Khaidiramsyah

NIM: 09121402058

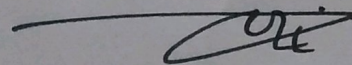
Palembang, 15 Juli 2019

Pembimbing I,



M. Fachrurrozi, M.T.
NIP. 198005222008121002

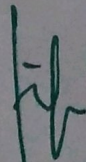
Pembimbing II,



Osvari Arsalan, M.T.
NIP. 198806282018031001

Mengetahui,

Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T.
NIP. 197706012009121004

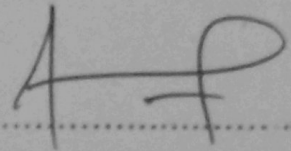
TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Selasa tanggal 15 Juli 2019 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Muhamad Khaidiramayah
NIM : 09121402058
Judul : Perbandingan Metode *Naive Bayes* dan Metode *KNN* untuk Klasifikasi Berita

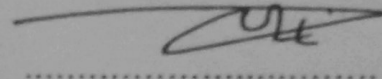
1. Ketua Penguji

M. Fachrurrozi, M.T.
NIP. 198005222008121002



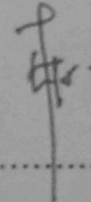
2. Sekretaris

Osvari Arsalan, M.T.
NIP. 198806282018031001



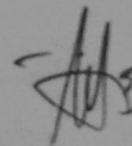
3. Penguji I

Yoppy Sazaki, M.T.
NIP. 197406062015109101

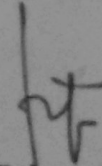


4. Penguji II

Muhammad Ali Buchari, M.T.
NIP. 198803302019031007



Mengetahui,
Ketua Jurusan Teknik Informatika


Rifkie Primartha, M.T.
NIP. 197706012009121004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Muhamad Khaidiramasyah
NIM : 09121402058
Program Studi : Teknik Informatika
Judul Skripsi : Perbandingan Metode *Naïve Bayes* dengan Metode
K-NN untuk Klasifikasi Berita
Hasil Pengecekan Software *iThenticate/Turnitin* : 10 %

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 15 Juli 2019



(Muhamad Khaidiramasyah)
NIM. 09121402058

Motto

- *Awali setiap kegiatan dengan sholat.*
- *Saat kesusahan di sanalah Allah menguji karena pada dasarnya Allah tidak akan menguji manusia di luar kemampuannya(QS. Al-Baqarah:286).*
- *Selalu ikhlas dan berlapang dada dalam bekerja, karena sesungguhnya Allah tidak suka dengan orang sombong lagi membanggakan diri(QS. Luqman:18).*
- *Tidak ada kebaikan yang sia-sia, Allah akan membalas perbuatan itu meskipun hanya sebesar dzarrah(QS. As4z-Zalzalah:7).*

Kupersembahkan karya tulis ini kepada :

- Allah SWT beserta kekasihnya yang kurindukan, baginda Rasulullah SAW.
- Ayah, Ibu, kakak dan adiku yang selalu mengingatkan.
- Dosen-dosen pengajarku yang selalu sabar dan membimbing dengan ikhlas.
- Teman-teman seperjuangan yang selalu ada untuk membantu dalam pembuatan skripsi, tidak bisa disebut satu persatu, tetapi saya ucapkan banyak terima kasih atas bantuannya

Perbandingan Metode Naïve Bayes dengan Metode K-NN untuk Klasifikasi Berita

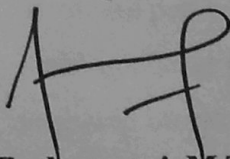
Oleh :
Muhamad Khaidiramasyah
09121402058

ABSTRAK

Berita merupakan informasi yang penting untuk diketahui, banyaknya jenis berita terkadang membuat pembaca menjadi bingung untuk memilih berita apa yang ingin dibaca. salah satu cara untuk memudahkan pembaca memilih berita secara otomatis dan akurat adalah dengan klasifikasi dokumen. Terdapat banyak metode klasifikasi yang dapat digunakan. Penelitian ini menggunakan metode *Naïve Bayes* dan Metode *K-Nearest Neighbor*, kedua metode ini dipilih karena metode *Naïve Bayes* adalah klasifikasi berdasarkan probabilitas, metode *K-Nearest Neighbor* adalah metode pembelajaran berdasarkan jarak terdekat dengan objek tersebut. Klasifikasi ini menggunakan data training sebanyak 1000 data dan data testing sebanyak 100 data. Hasil penelitian menunjukkan akurasi *K-Nearest Neighbor* lebih tinggi dengan $K=7$ akurasi didapat oleh *K-Nearest Neighbor* 97% dan *Naïve Bayes* mendapatkan akurasi sebesar 72%.

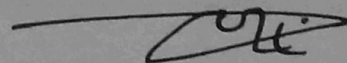
Kata Kunci: Klasifikasi, Metode K-Nearest Neighbor (K-NN), Metode Naive Bayes.

Pembimbing I,



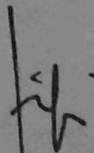
M. Fachrurrozi, M.T
NIP. 198005222008121002

Palembang, 15 Juli 2019
Pembimbing II,



Osvari Arsalan, M.T
NIP. 198806282018031001

Mengetahui,
Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T
NIP. 197706012009121004

KATA PENGANTAR



Alhamdulillah, puji syukur kehadiran Allah SWT atas segala nikmat, rahmat, dan karunia-Nya, sehingga penulis dapat menyelesaikan tugas akhir dengan judul “**Perbandingan Metode Naïve Bayes dengan Metode K-NN untuk Klasifikasi Berita**”. Tugas akhir ini disusun sebagai salah satu syarat kelulusan tingkat sarjana pada Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Penulis menyadari bahwa dalam penulisan tugas akhir ini tidak terlepas dari dukungan, bimbingan, kerja sama, dan motivasi dari berbagai pihak. Untuk selanjutnya penyusun mengucapkan banyak terima kasih kepada pihak-pihak yang telah membantu dalam penyelesaian tugas akhir ini, yaitu:

1. Orang tuaku M. Muslim dan Nitalia Mihardyanti yang senantiasa dalam memberikan kesabaran juga terus memberikan doa, nasihat, dan dukungan dari segala macam aspek yang begitu besar bagi penulis selama menempuh jenjang pendidikan juga teruntuk kakak dan adikku yang senantiasa hadir dalam membantu pembuatan tugas akhir ini;
2. Keluarga besar yang selalu mendoakan dan memberikan dukungan baik secara langsung maupun tidak langsung;
3. Bapak Jaidan Jauhari, S.Pd., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya;
4. Bapak Rifkie Primartha, M.T. selaku Ketua Jurusan Teknik Informatika Universitas Sriwijaya sekaligus sebagai Pembimbing Akademik bagi penulis selama masa perkuliahan;
5. Bapak M. Fachrurrozi, M.T. dan Kak Osvari Arsalan, M.T. selaku dosen pembimbing tugas akhir yang telah memberikan bimbingan, ilmu pengetahuan, dukungan, kritik dan saran yang membangun kepada penulis dalam proses penyusunan tugas akhir ini;

6. Bapak Yoppy Sazaki, M.T. dan Kak Muhammad Ali Buchari, M.T. selaku penguji yang telah memberikan kritik dan saran kepada penulis dalam penyempurnaan tugas akhir ini;
7. Mbak Titi beserta penerus tahtanya sebagai admin jurusan Teknik Informatika Bilingual yakni mbak Wiwin yang telah banyak membantu dan membimbing penulis dalam segala urusan administrasi juga dalam hal penjadwalan ujian yang luar biasa sekali;
8. Teman-teman dan kakak senior yang ikut serta membatu penulisan yang tidak bisa disebut satu persatu, tetapi penulis sangat berterima kasih kepada mereka semua;
9. Mereka yang telah menjadi *support* baik itu dalam segi strategi bertemu dengan pembimbing, strategi revisi, pembimbing berkas serta yang sama-sama berjuang dalam memperoleh gelar sarjananya.
10. Seluruh pihak yang turut membantu dalam penyusunan dan penyempurnaan tugas akhir ini yang tidak dapat penulis sebutkan satu-persatu.

Palembang, Juli 2019

Penulis.

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN TUGAS AKHIR	ii
TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xv
DAFTAR RUMUS	xvi

BAB I PENDAHULUAN

1.1 Latar Belakang Masalah	I-1
1.2 Perumusan Masalah	I-3
1.3 Tujuan Penelitian	I-3
1.4 Manfaat Penelitian	I-4
1.5 Batasan Masalah	I-4

BAB II TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu	II-1
2.2 Klasifikasi	II-2
2.3 <i>Preprocessing</i>	II-2
2.3.1 Tokenisasi	II-3
2.3.2 <i>Filtering</i>	II-3
2.3.3 <i>Stemming</i>	II-3
2.3.4 <i>Information Retrieval</i> (Temu Kembali Informasi).....	II-5

2.3.4.1	<i>Term Frequency</i>	II-6
2.3.4.2	<i>Inverse Document Frequency</i>	II-6
2.3.4.3	Metode Pembobotan TF – IDF	II-6
2.4	<i>Naive Bayes</i>	II-8
2.5	K-Nearest Neighbor	II-10
2.6	<i>Rational Unified Process (RUP)</i>	II-11

BAB III METODOLOGI PENELITIAN

3.1	Unit Penelitian	III-1
3.2	Metode Pengumpulan Data	III-1
3.2.1	Jenis Data	III-1
3.2.2	Sumber Data	III-1
3.2.3	Teknik Pengumpulan Data	III-1
3.3	Tahapan Penelitian	III-2
3.4	Diagram Blok Proses Umum Perangkat Lunak	III-2
3.5	Metode Pengembangan Perangkat lunak	III-5
3.6	Teknik Pengujian	III-7
3.7	Penjadwalan Tugas Akhir	III-8

BAB IV PENGEMBANGAN PERANGKAT LUNAK

4.1	Pendahuluan	IV-1
4.2.1	Pemodelan Bisnis	IV-1
4.2.2	Kebutuhan	IV-2
4.3	Analisis dan Desain	IV-4
4.3.1	Analisis Kebutuhan	IV-4
4.3.2	Analisis Data	IV-5
4.3.3	Analisis Metode <i>K-Nearest Neighbor</i>	IV-5
4.3.4	Analisis Metode <i>Naive Bayes</i>	IV-6
4.3.5	Implementasi	IV-6
4.4	Desain Perangkat Lunak	IV-7

4.5. Diagram Kelas Analisis.....	IV-10
4.6 <i>Sequence Diagram</i>	IV-14
4.7 Diagram Kelas.....	IV-17
4.8 Rancangan Antarmuka.....	IV-19
4.9 Implementasi Kelas	IV-20
4.10 Rencana Pengujian	IV-22
4.11 Implementasi.....	IV-24

BAB V HASIL DAN ANALISIS PENELITIAN

5.1 Pendahuluan	V-1
5.2 Data Hasil Penelitian.....	V-1
5.3 Analisis Hasil Pengujian	V-7

BAB VI KESIMPULAN DAN SARAN

6.1 Kesimpulan	VI-1
6.2 Saran.....	VI-1

DAFTAR PUSTAKA	xvi
----------------------	-----

DAFTAR GAMBAR

	Halaman
II-1. Gambar Struktur Proses RUP	II-1
III-1. Gambar Diagram Blok Proses Umum Perangkat Lunak.....	III-3
IV-1 Gambar Diagram <i>Use Case</i>	IV-7
IV-2 Kelas Analisis Mengelompokkan Berita <i>K-Nearest Neighbor</i>	IV-12
IV-3 Kelas Analisis Mengelompokkan Berita <i>Naive Bayes</i>	IV-13
IV-4 Kelas Analisis Load File format xls	IV-14
IV-5 <i>Sequence</i> diagram Klasifikasi Berita <i>K-Nearest Neighbor</i>	IV-15
IV-6 <i>Sequence</i> Diagram Klasifikasi Berita Naive Bayes	IV-16
IV-7 <i>Sequence</i> Diagram Membuka File	IV-17
IV-8 Kelas Diagram	IV-18
IV-9. Rancangan Antarmuka <i>MainForm</i>	IV-19
IV-10 Antarmuka <i>MainForm</i> Perangkat Lunak.....	IV-24
IV-11 Antarmuka Hasil klasifikasi <i>K-Nearest Neighbor</i>	IV-24
IV-12 Antarmuka Hasil klasifikasi <i>Naive Bayes</i>	IV-12

DAFTAR TABEL

	Halaman
II-1 Sampel Data Uji <i>KNN</i>	II-12
II-2 Perhitungan TF-IDF	II-13
II-3 Bobot Kata	II-16
II-4 <i>Laplace Smoothing</i>	II-18
II-5 Term yang sama dengan data <i>testing</i>	II-19
II-6 Sampel Data Uji Naive Bayes.....	II-22
II-7 Jumlah Term Pada Masing-Masing Dokumen.....	II-23
II-8 <i>Euclidean Distance</i>	II-25
II-9 Jarak Dokumen Uji dengan Dokumen Latih.....	II-26
II-10 Urutan Kemiripan Dokumen.....	II-26
III-1. Kegiatan Pengembangan Perangkat Lunak Berdasarkan RUP.....	III-6
III-2. Blok Penjadwalan Penelitian	III-8
IV-1 Kebutuhan Fungsional.....	IV-3
IV-2 Kebutuhan Non Fungsional	IV-4
IV-3 Definisi Aktor	IV-8
IV-4 Definisi <i>Use Case</i>	IV-8
IV-5 <i>Use Case</i> Membuka File.....	IV-9
IV-6 Skenario <i>Use Case</i> Kalsifikasi Naive Bayes	IV-9
IV-7 Skenario <i>Use Case</i> Klasifikasi <i>K-Nearest Neighbor</i>	IV-10
IV-8 Daftar Implementasi Kelas	IV-20
IV-9 Rencana Pengujian <i>Use Case</i> Memasukkan Data	IV-22
IV-10 Rencana Pengujian <i>Use Case</i> Klasifikasi <i>K-Nearest Neighbor</i>	IV-23
IV-11 Rencana Pengujian <i>Use Case</i> Klasifikasi <i>Naive Bayes</i>	IV-23
IV-12 Pengujian <i>Use Case</i> Membuka File.....	IV-27
IV-13 Pengujian <i>Use Case</i> Klasifikasi <i>K-Nearest Neighbor</i>	IV-28
IV-14 Pengujian <i>Use Case</i> Klasifikasi <i>Naive Bayes</i>	IV-27
V-1 Hasil Pengujian <i>KNN</i>	V-1
V-2 Hasil Pengujian <i>Naive Bayes</i>	V-2
V-3 Hasil Pengujian <i>KNN</i> dan <i>Naive Bayes</i>	V-3

V-4 Hasil Waktu Menghitung Akurasi Klasifikasi *KNN* dan *Naive Bayes*V-4

DAFTAR RUMUS

	Halaman
II-1. Rumus TF-IDF	II-7
II-2 Rumus TF-IDF	II-7
II-3 Rumus TF-IDF	II-7
II-4 Rumus TF-IDF	II-7
II-5 Rumus TF-IDF	II-8
II-6 Rumus <i>Naïve Bayes</i>	II-8
II-7 Rumus <i>Naïve Bayes</i>	II-8
II-8 Rumus <i>Naïve Bayes</i>	II-9
II-9 Rumus <i>Naïve Bayes</i>	II-9
II-10 Rumus <i>K-Nearest Neighbor</i>	II-10
III-1 Rumus Akurasi	III-7

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini menjelaskan secara detail tentang latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan dokumen penelitian.

1.2 Latar Belakang Masalah

Klasifikasi teks adalah proses penemuan fungsi terhadap suatu kelas yang telah ditetapkan, bertujuan untuk memprediksi teks tersebut tergolong pada kelas apa. Klasifikasi terdapat beberapa tahapan diantaranya *pre-processing*, *indexing* dan *dimensionality reduction* (Vijayan *et al.* 2017).

Metode yang telah banyak digunakan dalam klasifikasi teks yaitu *Decision Tree Classifier*, *Rule Base Classifier*, *Naïve Bayes Classifiers*, *Proximity Based Classifier*, *Neural Network Classifier* and *K-Nearest Neighbor Classifier* dan lain-lain. Masing-masing metode memiliki kelemahan dan keuntungan yang membuat metode tersebut cocok untuk model tertentu (Vijayan *et al.* 2017).

K-Nearest Neighbor (KNN) adalah jenis teks klasifikasi *non-parametric* yang menggunakan pengukuran dengan melihat kemiripan (*dot product*, *cosine similarity*) untuk mengelompokkan dokumen. Hasil dari *training data* adalah pengetahuan yang digunakan pada dokumen uji di *testing data*, dengan

membandingkan *cosine similarity* pada setiap data yang telah dilatih di fase *training data*. Data uji dikelompokkan ke masing-masing kelas sesuai dengan metode *cosine similarity*.

Naïve Bayes adalah metode klasifikasi sederhana yang menggunakan probabilitas dan bekerja berdasarkan asumsi independensi bersyarat dari klasifikasi teks. Penggolongan atau pembagian kelas di *Naïve Bayes* berdasarkan probabilitas terbesar pada metode *Naïve Bayes*.

Berkembangnya media masa serta pesatnya ilmu pengetahuan berita menjadi sangat penting untuk diketahui oleh banyak orang bahkan berita bisa menjadi bahan acuan untuk seseorang melakukan aktivitas. Padatnya aktivitas terkadang membuat orang untuk sulit untuk membedakan berita apa yang dipaparkan oleh *website* berita *online*, terkadang berita yang disampaikan tidak sesuai dengan judul berita tersebut.

Berdasarkan uraian diatas penulis ingin melakukan penelitian mengklasifikasi berita serta mencari metode apa yang terbaik untuk mengukur akurasi dan kecepatan dari metode *Naive Bayes* dan *K-Nearest Neighbor*.

1.3 Rumusan Masalah

Berdasarkan uraian pada latar belakang, rumusan masalah pada penelitian ini adalah membandingkan kedua metode dari segi akurasi untuk klasifikasi berita. Metode yang akan dibandingkan adalah metode *Naive Bayes* dan metode *K-Nearest Neighbor*.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini yaitu:

Mengetahui akurasi dari metode *Naive Bayes* dan metode *K-Nearest Neighbor* dalam mengklasifikasi berita.

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

1. Sebagai bahan studi dan tambahan ilmu pengetahuan dalam hal klasifikasi teks atau dokumen bagi mahasiswa/mahasiswi Fakultas Ilmu Komputer terutama Departemen Teknik Informatika yang ingin melakukan penelitian selanjutnya.
2. Sebagai masukan dan usulan kepada pembaca dalam menentukan metode apa yang terbaik untuk klasifikasi berita.

1.6 Batasan Masalah

Batasan masalah dari penelitian ini adalah sebagai berikut:

1. Berita yang diambil hanya dalam bahasa Indonesia.
2. Data berupa dokumen yang berisi berita dan diambil dari Kompas.com, Liputan6.com, republika.co.id, dan detik.com.
3. Berita akan dikelompokkan ke dalam berita politik, berita ekonomi, berita kriminal, dan berita olahraga.

1.7 Sistemika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

BAB I. PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini akan dibahas dasar-dasar teori yang digunakan dalam penelitian, seperti definisi klasifikasi, *K-Nearest Neighbor*, *Naive Bayes* dan penelitian lain yang relevan dengan penelitian yang sedang dikembangkan.

BAB III. METODOLOGI PENELITIAN

Pada bab ini akan dibahas mengenai tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Pada bab ini akan dibahas mengenai perancangan, lingkungan implementasi, dan hasil pengujian perangkat lunak.

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini, hasil pengujian berdasarkan langkah-langkah yang telah direncanakan disajikan. Analisis diberikan sebagai basis dari kesimpulan yang diambil dalam penelitian ini

BAB VI. KESIMPULAN DAN SARAN

Pada bab ini berisi kesimpulan dari semua uraian pada bab-bab sebelumnya dan juga berisi saran-saran yang diharapkan berguna dalam penerapan pengembangan perangkat lunak ini.

1.8 Kesimpulan

Berdasarkan uraian di atas maka dapat disimpulkan dalam penelitian ini, peneliti akan mengembangkan perangkat lunak perbandingan akurasi K-Nearest Neighbor (KNN) dan Naive Bayes berdasarkan kelas berita.

DAFTAR PUSTAKA

- Agusta, I. (2009). Empowering Kinship to Counter Control on Family Discourse in Indonesia. *Journal of Asia Pacific Studies*, 1(1), 49–68.
- An, Y., Sun, S., & Wang, S. (2017). Naive Bayes classifiers for music emotion classification based on lyrics. *Proceedings - 16th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2017*, (1), 635–638. <https://doi.org/10.1109/ICIS.2017.7960070>
- Bhakre, S. K., & Bang, A. (2016). Emotion recognition on the basis of audio signal using Naive Bayes classifier. *2016 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2016*, 2363–2367. <https://doi.org/10.1109/ICACCI.2016.7732408>
- Brusilovsky, P., Ahn, J., & Rasmussen, E. (2010). Teaching information retrieval with web-based interactive visualization. *Journal of Education for Library and Information Science*, 51(3), 187–200.
- Chandra, N., Khatri, S. K., & Som, S. (2017). Anti Social Comment Classification based on kNN Algorithm, 348–354.
- Han J & Kamber M, “Data Mining: Concepts and Techniques”, Morgan Kaufmann Publishers, Second Edition, 2006
- Harisinghaney, A., Dixit, A., Gupta, S., & Arora, A. (2014). Text and image based spam email classification using KNN, Naïve Bayes and Reverse DBSCAN algorithm. *ICROIT 2014 - Proceedings of the 2014 International Conference on Reliability, Optimization and Information Technology*, 153–155. <https://doi.org/10.1109/ICROIT.2014.6798302>
- Kowalski, Gerald, Mark T Maybury: *Information Retrieval Systems: Theory and Implementation*, Kluwer Academic Publishers 2011.
- Kruchten Philippe, "The Rational Unified Process: An Introduction", Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, Second Edition, 2000
- Naik, K., Dwarkadas, J., Joshi, A., Dwarkadas, J., Khanna, P., Shekokar, N., & Dwarkadas, J. (2017). A model to analyse social media data to gain a competitive edge. <https://doi.org/10.1109/ICCUBEA.2017.8463792>
- Netti, K., & Radhika, Y. (2016). A novel method for minimizing loss of accuracy in Naive Bayes classifier. *2015 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2015*, 1–4. <https://doi.org/10.1109/ICCIC.2015.7435801>
- Nurzahputra, A., & Muslim, M. A. (2016). Analisis Sentimen pada Opini Mahasiswa Menggunakan Natural Language Processing. *Seminar Nasional Ilmu Komputer, (Snik)*, 114–118.

- Salton, G. (1971). The Smart environment for retrieval system evaluation- advantages and problem areas, 316–329.
- Vijayan, V. K., Bindu, K. R., & Parameswaran, L. (2017). A Comprehensive Study of Text Classification Algorithms. *Ijsrcseit*, 4(1), 1109–1113.
- Wang, K., & Shang, W. (2017). Outcome prediction of DOTA2 based on Naïve Bayes classifier. *Proceedings - 16th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2017*, (1994), 591–593.
<https://doi.org/10.1109/ICIS.2017.7960061>
- Zainuddin, A. Z. A., Lee, K. Y., Mansor, W., & Mahmoodin, Z. (2017). Optimized KNN classify rule for EEG based differentiation between capable dyslexic and normal children. *IECBES 2016 - IEEE-EMBS Conference on Biomedical Engineering and Sciences*, 685–688.
<https://doi.org/10.1109/IECBES.2016.7843537>