

**KLASIFIKASI TEKS SENTIMEN NEGATIF KALIMAT BAHASA
INDONESIA PADA KOLOM KOMENTAR INSTAGRAM MENGGUNAKAN
ALGORITMA *NAÏVE BAYES* DENGAN *FEATURE EXTRACTION N-GRAM*
DAN TF-IDF**

TUGAS AKHIR



**UNIVERSITAS
BAKRIE**

ELSISMONE UTARI FITRI

1162001015

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER**

UNIVERSITAS BAKRIE

JAKARTA

2020

**KLASIFIKASI TEKS SENTIMEN NEGATIF KALIMAT BAHASA
INDONESIA PADA KOLOM KOMENTAR INSTAGRAM MENGGUNAKAN
ALGORITMA *NAÏVE BAYES* DENGAN *FEATURE EXTRACTION N-GRAM*
DAN TF-IDF**

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer**



Elsismone Utari Fitri

1162001015

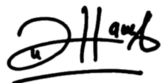
**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS BAKRIE
JAKARTA
2020**

HALAMAN PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar.

Nama : Elismone Utari Fitri

NIM : 1162001015

Tanda Tangan : 

Tanggal : 21 Agustus 2020





HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh:

Nama	Elsismone Utari Fitri
NIM	1162001015
Program Studi	Informatika
Fakultas	Teknik dan Ilmu Komputer
Judul Tugas Akhir	Klasifikasi Teks Sentimen Negatif Kalimat Bahasa Indonesia Pada Kolom Komentar Instagram Menggunakan Algoritma <i>Naive Bayes</i> dengan <i>Feature Extraction N-Gram</i> dan TF-IDF

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik dan Ilmu Komputer, Universitas Bakrie

DEWAN PEMBAHAS

Pembimbing 1	Prof. Dr. Hoga Saragih, S.T., M.T.	()
Pembimbing °	Ihsan Ibrahim, S.T., M.T.	()
Penguji I	Arief B. Suharko, BSEE, MSEE, Ph.D	()
Penguji 2	Ir. Magit Fitroni, M. Kom	()

Ditetapkan di	Jakarta
Tanggal	21 Agustus 2020

UCAPAN TERIMA KASIH

Alhamdulillahirabbil'alamin puji dan syukur atas kehadiran Allah SWT karena berkat segala rahmat dan karunia-NYA Tugas Akhir yang berjudul Klasifikasi Teks Sentimen Negatif Kalimat Bahasa Indonesia Pada Kolom Komentar Instagram Menggunakan Algoritma *Naïve Bayes* Dengan *Feature Extraction N-Gram* Dan TF-IDF dapat diselesaikan. Penyusunan Tugas Akhir ini tidak lepas dari berbagai kesulitan, rintangan dan hambatan dari awal hingga akhir penyusunan. Penulis mengucapkan terima kasih sebesar-besarnya kepada dosen pembimbing bapak Ihsan Ibrahim, S.T., M.T dan bapak Prof. Dr. Hoga Saragih, S.T., M.T yang telah membantu proses pengerjaan dan penyusunan Tugas Akhir ini dan yang telah meluangkan waktu serta mengarahkan pikiran dan tenaganya. Begitu banyak pihak yang turut dan membantu dan memberikan dukungan, masukan, nasihat, serta doa dalam penyusunan Tugas Akhir ini. Oleh karena itu, dengan segala hormat dan kerendahan hati, penulis mengungkapkan terima kasih kepada:

1. Papa, mama, dan adek yang tidak pernah lelah untuk selalu mengingatkan dan selalu memberikan doa, dukungan, motivasi, nasihat, semangat serta masukan.
2. Bapak Prof. Hoga Saragih, S.T., M.T. selaku Kepala Program Studi Informatika Bakrie yang telah membantu memberikan saran, masukan dan perbaikan dalam penelitian ini.
3. Bapak Ir. Magit Fitroni, M. Kom dan bapak Arief B. Suharko, BSEE, MSEE, Ph.D selaku dosen penguji yang telah memberikan perbaikan dan saran dalam penelitian ini.
4. *To my buddies and my special women in everything* Dinda dan Nida, terima kasih untuk semua hal yang telah kita hadapi dan jalani bersama selama masa perkuliahan. Perempuan-perempuan tangguh yang selalu memberikan semangat yang sangat luar biasa.
5. Arif Putra Aprideyarsa, *my bestfriend and my one call away* yang senantiasa dengan lapang dada menerima dan mau mendengarkan segala keluh kesah

- penulis dan selalu memberikan dukungan, doa, dan bantuan serta tidak pernah lelah untuk selalu memberikan semangat dan mengingatkan penulis.
6. Dezan Andhika dan Ilham Fawaz, teman diskusi dan pembimbing tambahan mandiri yang senantiasa membantu dan memberikan masukan dalam penelitian ini yang tidak pernah lelah mendengarkan ocehan penulis.
 7. Zikra Adiyofa, *my sister* and *my diary* yang selalu memberikan semangat dengan suara toanya dan selalu *stay with me* dalam kondisi apapun, dan Viny amelia arif yang selalu memebrikan dukungan.
 8. A Gregory Qonitah Michelle, Binya Amary, dan Refa Nurleana *my strong woman* yang selalu menginspirasi dan memotivasi penulis untuk harus selalu menjadi perempuan kuat dan tidak boleh lemah.
 9. Arun dan Luchia sahabat senasib, para pejuang perpunas, *my one call away* yang berjuang bersama dalam menyelesaikan penelitian ini.
 10. Cindy Agustia dan Fatimah yang selalu mengisi hari-hari penulis dengan semangat dan canda tawa selama proses penyusunan penelitian.
 11. Alya Priskalita dan Nabila Arza Utami yang selalu *support* dalam kondisi apapun, dua manusia yang selalu memberikan tawa.
 12. *My lovely* Syifa Awalita Arment yang senantiasa selalu sabar mendengarkan dan tidak pernah henti memberikan *support* dalam keadaan apapun, manusia yang menjadi *diary* penulis selama masa penyusunan penelitian.
 13. Kak Evie, Hani, Jogi, Dapot, Tamel *Squad* yang selalu membuat penulis tertawa dan menjadi semangat dalam mengerjakan penelitian
 14. Ermy Yuhari, Rizky, Syifa Nur Aini, Agust, Ode, Iki, Kharisma, Guntur, Tipa, Gita, Werin, dan Bang Ari yang telah memberikan dukungan dan banyak membantu selama masa perkuliahan.
 15. Bang Alim, Kak Amel TIF 14, Kak Rani, dan Bang Azmi yang telah memberikan dukungan, membantu, dan mengingatkan selama masa perkuliahan.
 16. Alif, Dinda, dan Nida, teman seperbimbingan yang sama-sama berjuang dalam menyelesaikan penelitian ini.

17. Teman-teman Informatika Angkatan 2016 yang telah memberikan dukungan, semangat, motivasi serta kebersamaan dalam suka maupun duka selama 4 tahun perkuliahan yang memberikan pengalaman dan pembelajaran selama masa perkuliahan di Universitas Bakrie
18. Kakak-kakak Angkatan 2015 dan adik-adik Angkatan 2017, 2018, dan 2019 Informatika Universitas Bakrie yang telah memberikan pengalaman, motivasi, dan semangatnya selama masa perkuliahan dan penyusunan Tugas Akhir ini.
19. Teman-teman Bakrie Angkatan 2016 yang tidak bisa disebutkan satu persatu namanya yang telah memberikan dukungan dalam penyusunan Tugas Akhir ini.
20. Seluruh Dosen Program Studi Informatika yang telah memberikan ilmu dan pembelajaran serta pengalaman yang sangat bermanfaat selama masa perkuliahan di Universitas Bakrie.
21. Mba Nur, Mas Anang, dan Mas Ali yang senantiasa sabar membantu dalam hal administrasi kampus.
22. Seluruh pihak Universitas Bakrie yang terlibat langsung maupun tidak yang telah memberikan bantuan, pengalaman, motivasi dan fasilitas yang sangat membantu selama masa perkuliahan dan penyusunan Tugas Akhir ini.

Semoga Allah SWT selalu memberikan keberkahan dan kebaikan kepada kita semua. Semoga Tugas Akhir ini dapat memberikan manfaat dan berguna bagi berbagai kalangan bidang pendidikan khususnya bidang Informatika.

Jakarta, 13 Agustus 2020

Penulis

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI

Sebagai sivitas akademik Universitas Bakrie, saya yang bertanda tangan di bawah ini:

Nama : Elsismone Utari Fitri
NIM : 1162001015
Program Studi : Informatika
Fakultas : Teknik dan Ilmu Komputer

Dengan pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bakrie Hak Bebas Royalti Noneksekutif (*Non-exclusive Royalti Free Right*) atas karya ilmiah saya yang berjudul:

Klasifikasi Teks Sentimen Negatif Kalimat Bahasa Indonesia Pada Kolom Komentar Instagram Menggunakan Algoritma *Naïve Bayes* Dengan *Feature Extraction N-Gram* dan *TF-IDF*. Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksekutif ini Universitas Bakrie berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*dataset*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/ pencipta dan sebagai pemilik Hak Cipta untuk kepentingan akademis.

Dengan demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di: Jakarta
Pada Tanggal: 21 Agustus 2020

Yang Menyatakan,



Elsismone Utari Fitri

**KLASIFIKASI TEKS SENTIMEN NEGATIF KALIMAT BAHASA
INDONESIA PADA KOLOM KOMENTAR INSTAGRAM MENGGUNAKAN
ALGORITMA NAÏVE BAYES DENGAN *FEATURE EXTRACTION N-GRAM*
DAN *TF-IDF***

Elsismone Utari Fitri

ABSTRAK

Saat ini Instagram berkembang dengan pesat. Berdasarkan hasil riset Wearesosial Hootsuite pengguna aktif Instagram di Indonesia tahun 2019 mencapai 55 juta pengguna. Instagram memberikan fitur komentar yang dapat digunakan setiap penggunanya. Meningkatnya komentar-komentar yang diberikan terutama komentar yang bersifat negatif menjadi masalah serius bagi kehidupan sosial. Hujatan merupakan suatu hal yang diucapkan tanpa adanya bukti yang benar. Sebagian pengguna media sosial tidak menyadari adanya hukum yang membahas tentang pencemaran nama baik atau sentimen negatif dan belum menyadari mana saja teks yang mengandung unsur tuduhan. Oleh karena itu, penulis mengusulkan penelitian klasifikasi teks sentimen negatif kalimat bahasa Indonesia dengan menggunakan metode *machine learning*. Klasifikasi teks sentimen negatif bahasa Indonesia ini menggunakan algoritma *Naïve Bayes Classification* (NBC) dengan 2 (dua) metode pendekatan *Feature Extraction* yaitu *N-gram Features* dan *Term Frequency Inverse Document Frequency* (TF-IDF), dengan menggunakan 500 data yang terdiri dari 400 data *training* dan 100 data *testing* pada 5 nama dengan masing-masing nama memiliki 20 komentar. Hasil pengukuran performa pada penelitian klasifikasi teks fitnah bahasa Indonesia dengan metode *confusion matrix 2x2* dengan hasil tingkat *accuracy* 87%, *precision* 100%, *recall* 85%, dan *f-measure* 92%.

Kata Kunci: Sentimen Negatif, *Machine Learning*, *Naïve Bayes Classification* (NBC), *N-gram Features*, *Term Frequency Inverse Document Frequency* (TF-IDF).

**CLASSIFICATION OF INDONESIAN SENTENCES TEXT SENTIMENT
NEGATIVE ON INSTRAGRAM COMMENTS USING NAÏVE BAYES
ALGORITHM WITH FEATURE EXTRACTION N-GRAM AND TF-IDF**

Elsismone Utari Fitri

ABSTRACT

Currently Instagram is growing rapidly. Based on the result of research on Wearesocial Hootsuite, active Instagram users in Indonesia in 2019 have reached 55 million users. Instagram provides a comment feature that can be used by every user. The increase in comments given to negative comments is a serious problem for social life. Sentiment negative is something that is said without true evidence. Some media users are not aware of the existence of laws that address defamation or blasphemy and not knowing which texts contain elements of blasphemy. Therefore, based on research on the classification of Indonesian sentiment negative text sentences using machine learning methods. Indonesian sentiment negative text classification uses the Naïve Bayes Classification (NBC) algorithm with 2 (two) Features Exreaction approaches, there are the N-gram Feature and the Term Frequency Inverse Document Frequency (TF-IDF), using 500 data consist of 400 training data and 100 tests data of 5 (five) name with each name having 20 comments. The results of the performance measurement in the Indonesian slander text classification study using the 2x2 confusion matrix method with 87% accuracy, 100% precission, 85% recall, and 92% f-measure.

Keywords: Sentiment Negative, Machine Learning, Naïve Bayes Classification (NBC), N-gram Features, Term Frequency Inverse Document Frequency (TF-IDF).

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	i
HALAMAN PENGESAHAN.....	ii
UCAPAN TERIMA KASIH	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI.....	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR RUMUS	xv
DAFTAR LAMPIRAN.....	xvi
DAFTAR SINGKATAN.....	xvii
DAFTAR ISTILAH	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan dan Manfaat Penelitian.....	4
1.3.1 Tujuan Penelitian	4
1.3.2 Manfaat Penelitian	4
1.4 Ruang Lingkup Penelitian	4
1.5 Sistematika Penelitian	5
BAB II <i>TEXT MINING</i> UNTUK KLASIFIKASI	6
2.1 Penelitian Terkait	6
2.2 <i>Machine Learning</i>	11
2.3 <i>Text Mining</i>	12
2.4 <i>Text Classification</i>	13
2.5 <i>Feature Extraction</i>	16
2.5.1 <i>N-gram</i>	16

2.5.2 <i>Term Frequency Inverse Document Frequency (TF-IDF)</i>	17
2.6 <i>Naïve Bayes Classification</i>	19
2.7 <i>Confusion Matrix</i>	21
BAB III DESAIN DAN PERANCANGAN	24
3.1 Pengumpulan Data	25
3.2 Pelabelan Data.....	26
3.3 <i>Preprocessing</i>	27
3.3.1 <i>Cleansing</i>	28
3.3.2 <i>Case Folding</i>	29
3.3.3 <i>Stemming</i>	30
3.3.4 <i>Stopword</i>	31
3.4 <i>Feature Extraction</i>	32
3.4.1 <i>N-gram</i>	32
3.4.2 <i>Term Frequency Inverse Document Frequency (TF-IDF)</i>	33
3.5 Pembagian Data.....	38
3.5.1 <i>Data Training</i>	39
3.5.2 <i>Data Testing</i>	39
3.6 Klasifikasi Algoritma <i>Naïve Bayes</i>	39
3.6.1 <i>Proses Data Training (Latih)</i>	41
3.6.2 <i>Proses Data Testing (Uji)</i>	44
3.7 Evaluasi dan Hasil Klasifikasi.....	46
BAB IV IMPLEMENTASI DAN HASIL PENELITIAN	47
4.1 Pengumpulan data	47
4.2 Pelabelan Data.....	48
4.3 <i>Preprocessing</i> Data	49
4.3.1 <i>Cleansing</i>	49
4.3.2 <i>Case Folding</i>	50
4.3.3 <i>Stemming</i>	51
4.3.4 <i>Stopword</i>	53
4.4 <i>Feature Extraction</i>	54

4.4.1 <i>N-Gram Features</i>	55
4.4.1.1 <i>Unigram</i>	55
4.4.1.2 <i>Bigram</i>	56
4.4.1.2 <i>Trigram</i>	57
4.4.2 <i>Term Frequency Inverse Frequency Document (TF-IDF)</i>	58
4.4.2.1 Menghitung Nilai <i>Term Frequency (TF)</i>	59
4.4.2.2 Menghitung Nilai <i>Inverse Document Frequency (IDF)</i>	60
4.4.2.3 Menghitung Nilai <i>TF-IDF</i>	60
4.5 <i>Klasifikasi Naïve Bayes</i>	62
4.5.1 <i>Proses Data Training (Latih)</i>	62
4.5.2 <i>Proses Data Testing (Uji)</i>	63
4.6 <i>Confusion Matrix</i>	65
BAB V SARAN DAN KESIMPULAN	69
5.1 <i>Kesimpulan</i>	69
5.2 <i>Saran</i>	69
DAFTAR PUSTAKA	71

DAFTAR TABEL

Tabel 2.1 Penelitian terkait klasifikasi teks.....	9
Tabel 2.2 <i>Confusion matrix</i>	21
Tabel 3.1 Contoh penerapan proses <i>cleansing</i>	28
Tabel 3.2 Contoh penerapan proses <i>case folding</i>	29
Tabel 3.3 Contoh penerapan proses <i>stemming</i>	30
Tabel 3.4 Contoh penerapan proses <i>stopword</i>	32
Tabel 3.5 Contoh penerapan proses <i>N-gram</i>	33
Tabel 3.6 <i>Text input</i> TF-IDF.....	34
Tabel 3.7 Contoh penerapan <i>Term Frequency</i> (TF).....	35
Tabel 3.8 Contoh penerapan <i>Document Frequency</i> (DF).....	36
Tabel 3.9 Contoh penerapan <i>Inverse Document Frequency</i> (IDF).....	37
Tabel 3.10 Contoh penerapan TF-IDF.....	38
Tabel 3.11 Contoh pelabelan data <i>training</i>	41
Tabel 3.12 Frekuensi kemunculan kata unik.....	42
Tabel 3.13 Model probabilstik.....	43
Tabel 3.14 Data <i>testing</i>	45
Tabel 3.15 Frekuensi kemunculan kata.....	45
Tabel 3.16 Menghitung <i>conditional probability</i>	45
Tabel 4.1 <i>Dataset</i>	49
Tabel 4.2 Data <i>testing</i>	62
Tabel 4.3 Perbandingan data <i>predicated</i> dan <i>actual</i>	65
Tabel 4.4 Nilai <i>confusion matrix</i>	67
Tabel 4.5 <i>Implementation result</i>	68
Tabel 1. Potongan data <i>testing</i>	76
Tabel 2. Frekuensi kemunculan kata unik.....	78
Tabel 3. Model probabilsitik.....	79
Tabel 4. Frekuensi kemunculan kata <i>testing</i>	80
Tabel 5. Menghitung <i>conditional probability testing</i>	80

DAFTAR GAMBAR

Gambar 2.1 <i>Machine learning techniques</i>	11
Gambar 2.2 Taksonomi metode <i>machine learning</i>	14
Gambar 2.3 Taxonomi klasifikasi	14
Gambar 2.4 Pembelajaran mesin dengan klasifikasi terpandu.....	16
Gambar 3.1 Kerangka penelitian.....	24
Gambar 3.2 <i>Flowchart</i> sistem	25
Gambar 3.3 <i>Dataset</i>	26
Gambar 3.4 <i>Flowchart preprocessing</i>	27
Gambar 3.5 <i>Flowchart cleansing</i>	28
Gambar 3.6 <i>Flowchart case folding</i>	29
Gambar 3.7 <i>Flowchart stemming</i>	30
Gambar 3.8 <i>Flowchart stopword</i>	31
Gambar 3.9 <i>Flowchart N-gram</i>	32
Gambar 3.10 <i>Flowchart TF-IDF</i>	34
Gambar 3.11 <i>Flowchart Naïve Bayes Classification</i>	40
Gambar 4.1 Potongan data pada file datasetNew.csv	48
Gambar 4.2 <i>Dataset</i> penelitian.....	48
Gambar 4.3 <i>Source code cleansing</i>	50
Gambar 4.4 Potongan <i>output cleansing</i>	50
Gambar 4.5 <i>Source code case folding</i>	51
Gambar 4.6 Potongan <i>output case folding</i>	51
Gambar 4.7 <i>Source code stemming</i>	52
Gambar 4.8 Potongan <i>output stemming</i>	52
Gambar 4.9 Potongan <i>file kamus stopword</i>	53
Gambar 4.10 <i>Source code stopword</i>	54
Gambar 4.11 Potongan <i>output stopword</i>	54
Gambar 4.12 <i>Souce code unigram</i>	55

Gambar 4.13 Potongan <i>output unigram</i>	56
Gambar 4.14 <i>Source code bigram</i>	56
Gambar 4.15 Potongan <i>output bigram</i>	57
Gambar 4.16 <i>Source code trigram</i>	57
Gambar 4.17 Potongan <i>output trigram</i>	58
Gambar 4.18 Potongan kamus <i>query</i>	59
Gambar 4.19 <i>Source code TF</i>	59
Gambar 4.20 <i>Source code IDF</i>	60
Gambar 4.21 <i>Source code TF-IDF</i>	61
Gambar 4.22 Potongan <i>output TF-IDF</i>	61
Gambar 4.23 <i>Source code data training</i>	63
Gambar 4.24 <i>Source code data testing</i>	64
Gambar 4.25 Hasil klasifikasi <i>Naïve Bayes</i>	64
Gambar 4.26 <i>Source code confusion matrix</i>	66
Gambar 4.27 <i>Source code pengukuran peforma</i>	66
Gambar 4.28 <i>Output confusion matrix</i>	67
Gambar 1. Menghitung <i>prior</i>	76
Gambar 2. Menghitung <i>conditional probability</i>	76
Gambar 3. Mencari <i>Cmap</i>	80
Gambar 4. Penentuan kelas berdasarkan <i>Cmap</i>	81

DAFTAR RUMUS

Persamaan 2.1 <i>Term Frequency</i> (TF).....	17
Persamaan 2.2 <i>Inverse Document Frequency</i> (IDF).....	17
Persamaan 2.3 Persamaan TF-IDF.....	17
Persamaan 2.4 Persamaan <i>Teorema Bayes</i>	18
Persamaan 2.5 Penyederhaan Persamaan <i>Toerema Bayes</i>	19
Persamaan 2.6 Persamaan <i>Prior</i>	19
Persamaan 2.7 Persamaan <i>Posterior</i>	19
Persamaan 2.8 Perhitungan Probabilitas <i>Naïve Bayes</i>	19
Persamaan 2.9 Penyederhanaan Probabilitas <i>Naïve Bayes</i>	20
Persamaan 2.10 <i>Accuracy</i>	21
Persamaan 2.11 <i>Precision</i>	21
Persamaan 2.12 <i>Recall</i>	22
Persamaan 2.13 <i>F-measure</i>	22

DAFTAR LAMPIRAN

Lampiran 1. Klasifikasi teks	76
Lampiran 2. Potongan <i>Dataset</i>	82
Lampiran 3. <i>Stopword</i>	92
Lampiran 4. <i>Query</i>	101

DAFTAR SINGKATAN

TF	<i>Term Frequency</i>
IDF	<i>Inverse Document Frequency</i>
TF-IDF	<i>Term Frequency Inverse Document Frequency</i>
ML	<i>Machine Learning</i>
SVM	<i>Support Vector Machine</i>
NBC	<i>Naïve Bayes Classification</i>
TP	<i>True Positive</i>
FP	<i>False Positive</i>
TN	<i>True Negative</i>
FN	<i>False Negative</i>

DAFTAR ISTILAH

<i>Accuracy</i>	Jumlah dokumen yang diklasifikasi dengan benar.
<i>Artificial intelligence</i>	Kecerdasan yang ditambahkan kepada suatu sistem yang bisa diatur dalam konteks ilmiah.
<i>F-measure</i>	Digunakan untuk dan mendapatkan nilai kombinasi dari hasil perhitungan <i>precision</i> dan <i>recall</i> .
<i>Confusion matrix</i>	Metode yang digunakan dalam konsep data mining untuk menghitung nilai akurasi dan menganalisis hasil klasifikasi menggunakan <i>matrix</i> .
<i>Precision</i>	Jumlah dokumen yang relevan dengan hasil pemrosesan yang dicari.
<i>Posterior</i>	Peluang kategori teks atau kelas yang akan diklasifikasi terhadap kemunculan indeks kata.
<i>Preprocessing</i>	Proses pembersihan data dan mempersiapkan teks untuk diklasifikasi.
<i>Prior</i>	Peluang kemunculan kategori atau kelas yang akan diklasifikasi.
<i>Recall</i>	Digunakan untuk mengukur tingkat keberhasilan sistem dalam menghitung seberapa banyak dokumen yang berhasil terklasifikasi dari jumlah seluruh data yang digunakan.
<i>Supervised learning</i>	Metode pembelajaran mesin yang berkaitan dengan data terklasifikasi atau berlabel dan membutuhkan data <i>training</i> dan data <i>testing</i> .
<i>Text Mining</i>	Perluasan dari data <i>mining</i> yang dirancang untuk menangani data tidak terstruktur.