



YAYASAN PERGURUAN CIKINI  
INSTITUT SAINS DAN TEKNOLOGI NASIONAL

Jl. Moh. Kahfi II, Bhumi Srengseng Indah, Jagakarsa, Jakarta Selatan 12640  
Telp. 021-7270090 (hunting), Fax. 021-7866955, hp: 081291030024  
Email : humas@istn.ac.id Website : www.istn.ac.id

**SURAT PENUGASAN TENAGA PENDIDIK**  
**Nomor : 86/03.1-I/IX/2023**  
**SEMESTER GANJIL TAHUN AKADEMIK 2023/2024**

<b>Nama</b>	: Aryo Nur Utomo, S.T., M.Kom.	<b>Status Pegawai</b>	: Edukatif Tetap		
<b>NIP/NIK/NIDN</b>	: 01.121225/35091009/0319046803	<b>Program Studi</b>	: Sistem Informasi		
<b>Jabatan Akademik</b>	: Asisten Ahli				
Bidang	Perincian Kegiatan	Ruang/ Tempat	Jam/ Minggu	Kredit (sks)	Keterangan
I PENDIDIKAN DAN PENGAJARAN	<b>MENGAJAR DI KELAS (KULIAH/RESPONSI DAN LABORATORIUM)</b>				
	1. Dasar Pemrograman Komputer	A-6	2 Jam/Minggu	0	Senin, 08:00-09:40
	2. Pengelolaan Layanan IT	A-5	2 Jam/Minggu	1	Senin, 08:00-09:40
	3. Sistem Temu Kembali Informasi	A-3	2 Jam/Minggu	1	Senin, 14:30-16:30
	4. Analisa dan Perancangan Algoritma	A-6	2 Jam/Minggu	1	Senin, 10:30-12:10
	5. Pemrograman Jaringan	A-1	3 Jam/Minggu	1,5	Selasa, 14.41-15.40
	6. Komputer Forensik	A-2	3 Jam/Minggu	1,5	Selasa, 10:00-11:40
	7. Pembelajaran Mesin	A-1	3 Jam/Minggu	1,5	Rabu, 13.00-14.40
	8. Algoritma dan Pemrograman	A-3	3 Jam/Minggu	1,5	Kamis, 08:00-09:40
	9. Manajemen Proyek Perangkat Lunak	A-1	3 Jam/Minggu	1,5	Kamis, 10:00-11:40
	10. Bisnis Digital	A-2	2 Jam/Minggu	1	Jumat, 13:00-14:40
	11. Mobile Technology	A-3	2 Jam/Minggu	0	Senin, 10:00-11:40
	12. Pemrograman Berorientasi Objek	A-5	2 Jam/Minggu	0	Rabu, 08:00-09:40
	13. Keamanan sistem informasi	A-5	2 Jam/Minggu	0	Rabu, 15.30-17.00
	14. Menduduki Jabatan Struktural Ka.Prodi Tek. Informatika / Sistem Informasi		36 Jam/Minggu	3	Insidental
II PENELITIAN	Penulisan Karya Ilmiah			1	
II PENGABDIAN DAN MASYARAKAT	Pelatihan dan Penyuluhan			0	
IV UNSUR-UNSUR PENUNJANG	Berperan Serta Aktif dalam Pertemuan Ilmiah/Seminar			0	
	<b>Jumlah Total</b>			<b>16</b>	

Kepada yang bersangkutan akan diberikan gaji/honorarium sesuai dengan peraturan penggajian yang berlaku di Institut Sains dan Teknologi Nasional, penugasan ini berlaku tanggal 1 September 2023 sampai dengan 29 Februari 2024.

**Tembusan :**

1. Direktur Akademik - ISTN
2. Direktur Non Akademik - ISTN
3. Ka. Biro Sumber Daya Manusia - ISTN
4. Kepala Program Studi Sistem Informasi
5. Arsip

Jakarta, 1 September 2023  
Dekan

(Marhaeni, S.Kom., M.Kom.)



**DAFTAR HADIR PESERTA KULIAH MAHASISWA****GANJIL REGULER TAHUN 2023/2024**

FAK / JURUSAN : Sistem Informasi S1 HARI/TANGGAL : Senin  
MATA KULIAH : IT Service Management  
KELAS / PESERTA : A / 4 JAM KULIAH : 15.30-17.00  
KURIKULUM : 2023  
DOSEN : Aryo Nur Utomo, ST.M.Kom. RUANG :

NO	NIM	NAMA MAHASISWA	TANGGAL PERTEMUAN							JUMLAH	
			27/ 11	4/ 12	11/ 12	18/ 12	26/ 12	03/ 01	08/ 01		19/ 01
1	20350003	Anisa Qadri Kurniasih	√	√	√	√	√	√	√	U	7
2	20350004	Rizky Fauzi Ramadhan	√	√	√	√	√	√	√	U	7
3	20350006	Miftah Zaidan Falih	√	√	√	√	√	√	√	U	7
4	20350008	Muhammad Ibnu Afan Fuadi	√	√	√	√	√	√	√	U	7

Jakarta , ..... Februari 2024

Dosen Pengajar

( Aryo Nur Utomo, ST.M.Kom)



**BERITA ACARA PERKULIAHAN**  
**(PRESENTASI KEHADIRAN DOSEN)**  
**SEMESTER GANJIL TAHUN AKADEMIK 2023/2024**  
**PROGRAM STUDI SISTEM INFORMASI S1 FSTI-ISTN**

Mata Kuliah : Sistem Temu Kembali Informasi	Semester : 63571PSF01
Dosen : Aryo Nur Utomo, ST, M.Kom	SKS : 2
Hari : Senin	Kelas : A
Jam : 15:30-17:40	Ruang : A-2

No.	TANGGAL	MATERI KULIAH	JML MHS HADIR	TANDA TANGAN DOSEN
9.	27- November 2023	Text Classification	4	<i>Ah</i>
10.	4-Desember 2023	Text Clustering	4	<i>Ah</i>
11.	11- Desember 2023	Text Summarization	4	<i>Ah</i>
12.	18- Desember 2023	Question Answering System	4	<i>Ah</i>
13.	26- Desember 2023	Probabilistic in Information Retrieval	4	<i>Ah</i>
14.	3-Januari 2024	Cross Language in Information Retrieval	4	<i>Ah</i>
15.	8-Januari 2024	Web Search. Gathering information	4	<i>Ah</i>
16.	15-Januari 2024	<b>UJIAN AKHIR SEMESTER (UAS)</b>	4	<i>Ah</i>

DOSEN PENGAJAR

(Aryo Nur Utomo, ST, M.Kom)

# TEXT CLASSIFICATION

1

# DEFINISI

- Fungsi untuk mengklasifikasikan atau mengelompokkan teks (atau dokumen) ke dalam sebuah label tertentu
- Contoh:
  - Spam Filtering:
    - Mengelompokkan input (misalnya sms atau email) ke dalam kelompok spam atau bukan spam
  - News Clustering:
    - Mengelompokkan artikel berita ke dalam kategori tertentu seperti politik, olahraga, ekonomi, dst
- Dapat disebut sbg “Text Classification”, “Document Classification”, “Text Categorization”, “Document Categorization”

# APPLICATION OF TEXT CLASSIFICATION

- Spam filtering
- News clustering
- Sentiment analysis
- Plagiarism detection
- Essay scoring

# SPAM FILTERING

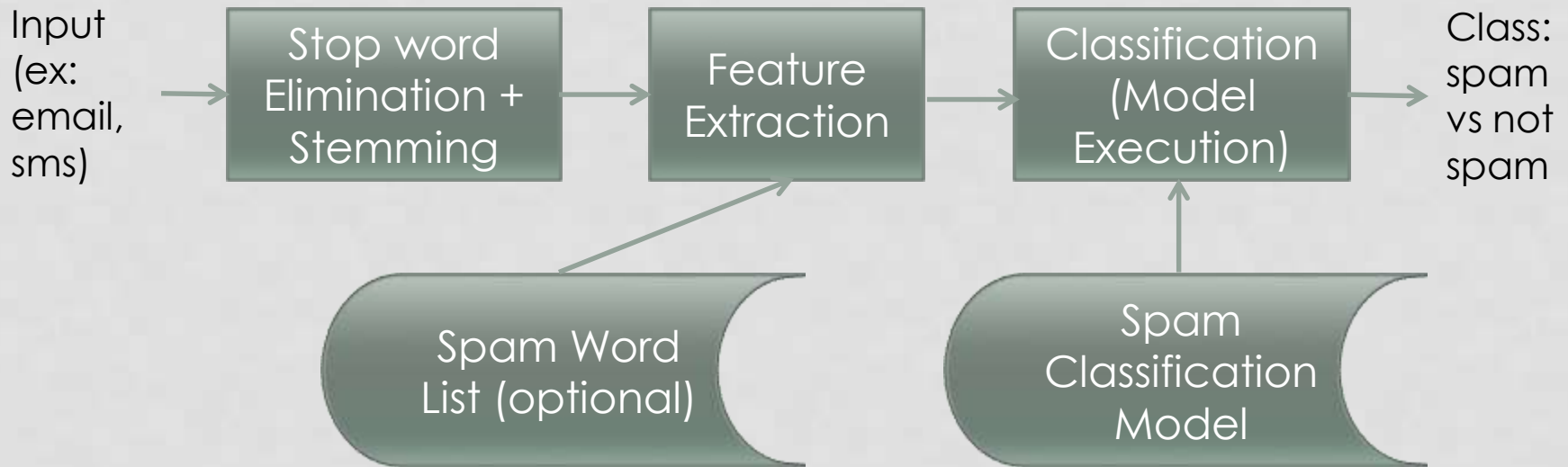
- Mengklasifikasikan email/sms ke dalam kelas spam dan bukan spam → 2 kelas
- Cara sederhana:
  - Menggunakan daftar kata berupa kata-kata yg banyak muncul di spam dan tetapkan aturan misalnya jika 75% kata-kata pada sms/email termasuk pada daftar kata tadi maka sms/email tsb adalah spam
  - Daftar kata → disebut **feature**
  - Aturan 75% → disebut **teknik klasifikasi** yang bisa berupa rule atau model

# PENDEKATAN UTAMA

- Terdiri atas **FEATURE** dan **TEKNIK KLASIFIKASI (RULE/MODEL)**
- Contoh Feature utk Spam Filtering:
  - Gunakan intuisi dalam menilai sebuah email/sms apakah berupa spam atau bukan
    - melihat kata-kata yang sering muncul pada email/sms spam → content based → NLP
    - melihat pola dari alamat pengirim
    - melihat pola dari subjek
- Teknik Klasifikasi:
  - Rule based (manual)
    - menggunakan aturan yang ditetapkan manusia
  - statistical based
    - menggunakan aturan yang dipelajari secara otomatis dari data (disebut juga training data) yang sudah ada dan dilabeli



# EXAMPLE OF STATISTICAL BASED SPAM FILTERING



# FEATURE

- Lexical based
  - Idea: to classify an input based on its words. To decrease the number of words, the words can be selected first, by using:
    - Specific word list (manually selected)
    - Named entity
    - TFxIDF
    - Mutual information
    - POS Tag information (for example: only Noun & Verb)
- Syntactical parser
  - Idea: to take into account, the word order or grammar; or to do word selection based on its syntactical information
  - Shallow parser
  - Deep parser
  - N-gram
- Semantic

# LEXICAL FEATURE UTK SPAM FILTERING

- Lexical based
  - Membuat daftar kata yang sering muncul pada spam → rule based
    - dibuat secara manual melalui pengamatan thd data spam
    - diberi nilai threshold
    - if “jumlah kata spam > konstanta” then spam
  - statistical based
    - penetapan aturan spam oleh algoritma (machine learning)
    - Contoh: if email mengandung kata  $w_1$  then spam ; dimana kata  $w_1$  diperoleh algoritma melalui data yang sudah dilabeli (training data)

# CONTOH LEXICAL FEATURE UTK SPAM FILTERING

- “Ini mama ... tolong kirim pulsa ke nomor hp ini”
- daftar kata: mama, tolong, kirim, pulsa, nomor
- Training data
  - fitur berupa jumlah kata pada email/sms yang termasuk daftar kata spam.
    - jumlah kata spam: **5** → prosentase kata spam: **0.25**
    - jumlah kata total: **20**
    - fitur kelas: **spam**
  - fitur berupa semua kata yang sering muncul di spam
    - mama: 1
    - tolong: 1
    - kirim: 1
    - pulsa: 1
    - nomor: 1

# DAFTAR KATA (LEXICAL BASED)

- kelemahan:
  - kata yang sering muncul pada spam, bisa jadi merupakan kata umum (*common words*) yang memang frekuensi kemunculannya banyak: ini, di, dst
    - stop word elimination
  - kata yang sering muncul pada spam, bisa jadi juga sering muncul pada not spam, meskipun bukan berupa kata umum : mama
    - pembobotan nilai kata
      - TF x IDF = frequency term / document frequency (yg mengandung term)
        - idf = 1/df
        - idf = log (N/df)
      - Mutual Information (MI) =  $MI(t, l) = \log_2 \frac{p(t, l)}{p(t) \times p(l)} \approx \log_2 \frac{A \times N}{(A + B) \times (A + C)}$

$N$  = Jumlah seluruh dokumen yang telah dilabeli

$A$  = Frekuensi kemunculan kata (t) dengan label (l)

$B$  = Frekuensi kemunculan kata (t) tanpa label (l)

$C$  = Jumlah dokumen dengan label (l) tanpa Term (t)

# TEKNIK KLASIFIKASI PADA TEXT CATEGORIZATION

- Clustering/Pengelompokan
  - Mengelompokkan dokumen dengan nilai fitur yang mirip
  - news clustering, email categorization
- Klasifikasi
  - Mengklasifikasikan dokumen ke dalam sebuah label tertentu (yg sudah didefinisikan thd setiap training data) berdasar nilai fiturnya
  - Spam filtering, Sentiment analysis, email categorization
- Comparison
  - Perbandingan dgn referensi
  - Essay scoring, plagiarism detection

# CONTOH KLASIFIKASI: NAIVE BAYES CLASSIFIER

- Teori bayes (likelihood)

$$p(C|D) = \frac{p(C)}{p(D)} p(D|C)$$

$$p(D|C) = \prod_i p(w_i|C)$$

# DAFTAR NILAI

## SEMESTER GANJIL REGULER TAHUN 2023/2024

Program Studi : Sistem Informasi S1  
Matakuliah : Sistem Temu Kembali Informasi  
Kelas / Peserta : A  
Perkuliahan : Kampus ISTN Bumi Srengseng Indah  
Dosen : Aryo Nur Utomo, ST.M.Kom.

Hal. 1/1

No	NIM	N A M A	ABSEN	TUGAS	UTS	UAS	MODEL	PRESENTASI	NA	HURUF
			10%	0%	45%	45%	0%	0%		
1	20350003	Anisa Qadri Kurniasih	100	0	88	84	0	0	87.4	A
2	20350004	Rizky Fauzi Ramadhan	100	0	75	71	0	0	75.7	A-
3	20350006	Miftah Zaidan Falih	100	0	88	84	0	0	87.4	A
4	20350008	Muhammad Ibnu Afan Fuadi	100	0	60	58	0	0	63.1	C+

Rekapitulasi Nilai							
A	2	B+	0	C+	1	D+	0
A-	1	B	0	C	0	D	0
		B-	0	C-	0	E	0

Jakarta,30 January 2024

Dosen Pengajar

**Aryo Nur Utomo, ST.M.Kom.**