



YAYASAN PERGURUAN CIKINI
INSTITUT SAINS DAN TEKNOLOGI NASIONAL
FAKULTAS SAINS DAN TEKNOLOGI INFORMASI

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

SURAT PENUGASAN TENAGA PENDIDIK

Nomor : 204 / 03.1 – I / IX / 2022

SEMESTER GANJIL, TAHUN AKADEMIK 2022 / 2023

Nama	: Aryo Nur Utomo, ST.,M.Kom.	Status Pegawai	: Edukatif Tetap		
NIK	: 01.121225	Program Studi	: Sistem Informasi		
Jabatan Akademik	: Asiste Ahli				
Bidang	Perincian Kegiatan	Ruang/ Tempat	Jam/ Minggu	Kredit (sks)	Hari / Waktu
I PENDIDIKAN Dan PENGAJARAN	MENGAJAR DI KELAS (KULIAH / RESPONSI DAN LABORATORIUM)				
	1. Cloud Computing (SI)	A-1	1 Jam/Minggu	1	Senin / 08:00-09:40
	2. Sistem Pendukung Keputusan (SI)	D-2	1,5 Jam/Minggu	1,5	Selasa / 08:00-10:00
	3. Sistem Temu Kembali Informasi(SI)	E-4	1 Jam/Minggu		Senin /15:30-17:00
	4. IT Service Management (SI)	D-3	1 Jam/Minggu	1	Jum'at / 14:30-16:00
	5. Keamanan Sistem Informasi (SI)	E-1	1 Jam/Minggu	1	Rabu / 15:30-17:00
	6. Algoritma dan Pemrograman (TIF)	A-2	1 Jam/Minggu	1	Kamis / 08:00-09:40
	7. Analisis dan Perancangan Algoritma (TIF)	A-1	1 Jam/Minggu	1	Senin / 10.30-12.10
	8. Pemrograman Jaringan (Java/Python)	D-2	1,5 Jam/Minggu	1,5	Selasa / 14.41-15.40
	9. Pembelajaran Mesin (TIF)	E-1	1,5 Jam/Minggu	1,5	Rabu / 13.00-14.40
	10. Komputer Forensik (TIF)	A-2	1,5 Jam/Minggu	1,5	Jumat / 12.30-14.00
	11. Pengelolaan Layanan TI (ITSM) (TIF)	A-1	1 Jam/Minggu	1	Senin / 08.00-09.40
	12. Data Compress & Coding (PIGS)	A-2	1 Jam/Minggu		Selasa / 08.00-09.40
	13. Manajemen Proyek Perangkat Lunak	A-3	1,5 Jam/Minggu		Kamis / 10.00-11.30
14. Menduduki Jabatan Struktural (Ka.Prodi TIF)			20 Jam/Minggu	3	
II PENELITIAN	Penulisan Karya Ilmiah			1	
III PENGABDIAN DAN MASYARAKAT	Pelatihan dan Penyuluhan				
IV UNSUR-UNSUR PENUNJANG	Berperan Serta Aktif dalam Pertemuan Ilmiah/Seminar				
Jumlah Total				16	

Kepada yang bersangkutan akan diberikan gaji / honorarium sesuai dengan peraturan penggajian yang berlaku di Institut Sains Dan Teknologi Nasional
Penugasan ini berlaku dari tanggal **02 September 2022** sampai dengan tanggal **29 Februari 2023**.

Jakarta, 30 September 2022
Dekah,

(Marnaeni, S.Kom.,M.Kom.)



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.

DAFTAR HADIR PESERTA KULIAH MAHASISWA**GANJIL REGULER TAHUN 2022/2023**

FAK / JURUSAN : Sistem Informasi S1 HARI/TANGGAL : Senin
MATA KULIAH : Sistem Temu Kembali Informasi
KELAS / PESERTA : A / 2 JAM KULIAH : 15.30-17.00
KURIKULUM : 2018
DOSEN : Aryo Nur Utomo, ST.M.Kom. RUANG :

NO	NIM	NAMA MAHASISWA	TANGGAL PERTEMUAN								JUMLAH
			21/ 11	28/ 11	05/ 12	12/ 12	19/ 12	26/ 12	02/ 01	16/ 01	
1	16350021	Inggar Ananto	√	√	√	√	√	√	√	U	7
2	20350501	Achmad Zikri	√	√	√	√	√	√	√	U	7

Jakarta , Februari 2023

Dosen Pengajar

(Aryo Nur Utomo, ST.M.Kom)



BERITA ACARA PERKULIAHAN
(PRESENTASI KEHADIRAN DOSEN)
SEMESTER GANJIL TAHUN AKADEMIK 2022/2023
PROGRAM STUDI SISTEM INFORMASI S1 FSTI-ISTN

Mata Kuliah : Sistem Temu Kembali Informasi	Semester : 357001
Dosen : Aryo Nur Utomo, ST, M.Kom	SKS : 2
Hari : Senin	Kelas : A
Jam : 15:30-17:40	Ruang : B-2

No.	TANGGAL	MATERI KULIAH	JML MHS HADIR	TANDA TANGAN DOSEN
9.	21-November 2022	Text Classification	2	Ah
10.	28-November 2022	Text Clustering	2	Ah
11.	5-Desember 2022	Text Summarization	2	Ah
12.	12-Desember 2022	Question Answering System	2	Ah
13.	19-Desember 2022	Probabilistic in Information Retrieval	2	Ah
14.	26-Desember 2022	Cross Language in Information Retrieval	2	Ah
15.	2-Januari 2023	Web Search. Gathering information	2	Ah
16.	16-Januari 2023	UJIAN AKHIR SEMESTER (UAS)	2	Ah

DOSEN PENGAJAR

(Aryo Nur Utomo, ST, M.Kom)

DAFTAR NILAI

SEMESTER GANJIL REGULER TAHUN 2022/2023

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
			0%	0%	50%	50%	0%	0%		
1	16350021	Inggar Ananto	100	0	86	84	0	0	85	A
2	20350501	Achmad Zikri	100	0	80	81	0	0	80.5	A

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

Jakarta, 18 February 2023

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

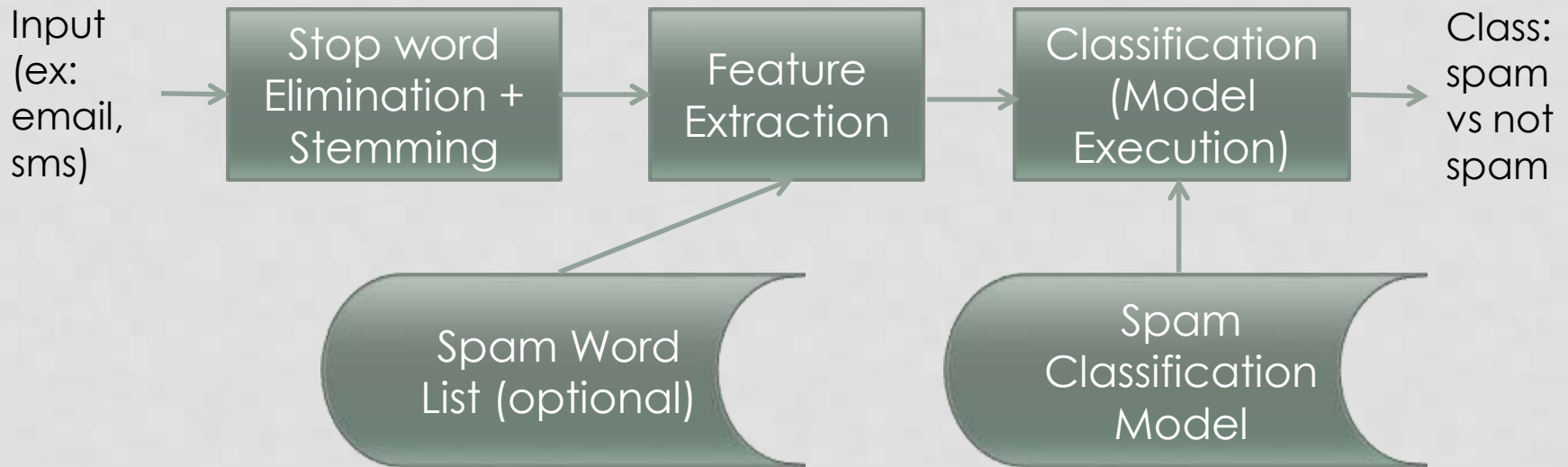
SPAM FILTERING

- 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)

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

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