	Universitas Negeri Surabaya S1 Sains Data	Kode Dokumen

RENCANA PEMBELAJARAN SEMESTER

MATA KULIAH (MK)	KODE	Rumpun MK	BOBOT (sks)	SEMESTER	Tgl Penyusunan
Pemrosesan Teks	4920202030		2		25 November 2026

OTORISASI	Pengembang RPS	Koordinator RMK	Koordinator Program Studi
	Nurul Laili, S.Pd., M.Sc.	Dr. Atik Wintarti, M.Kom	YULIANI PUJI ASTUTI

Deskripsi Singkat MK	Mata kuliah ini merupakan mata kuliah berbasis proyek yang mempelajari tentang teknik dasar untuk mengolah data teks. Pada mata kuliah akan diperkenalkan konsep morfologi bahasa, representasi teks, pre-processing, ekstraksi fitur hingga mendapatkan informasi seperi similarity dan klasterisasi teks. Topik-topik yang dibahas meliputi: morfologi bahasa, represntasi string, regex, tokenisasi, text pre-processing, Bag of Words, TF-IDF, word similarity, word clustering, dan web scraping. Mahasiswa akan membuat proyek berkelompok untuk mengaplikasikan teori dan konsep pemrosesan teks pada permasalahan di bidang Sains Data.

Capaian Pembelajaran (CP)

CPL-PRODI yang dibebankan pada MK

CPL-3	Mengembangkan pemikiran logis, kritis, sistematis, dan kreatif dalam melakukan pekerjaan yang spesifik di bidang keahliannya serta sesuai dengan standar kompetensi kerja bidang yang bersangkutan
CPL-19	Mampu menerapkan pemikiran logis, kritis, sistematis, dan inovatif dalam mendesain, mengimplementasi dan mengevaluasi ilmu pengetahuan dan teknologi yang memperhatikan dan menerapkan nilai humaniora dalam bidang sains data
CPL-22	Mampu merancang dan mengembangkan algoritma untuk berbagai keperluan seperti analisis big data, kecerdasan artifisial, basis data, penambangan data, statistika inferensial, desain dan analisis algoritma, dan data warehouse
CPL-26	Menguasai konsep teoritis bidang pengetahuan sains data khusus dalam bidang pengetahuan tersebut secara mendalam, serta mampu memformulasikan penyelesaian masalah prosedural

Capaian Pembelajaran Mata Kuliah (CPMK)

CPMK-1	Mampu merancang penyelesaian masalah pada data teks menggunakan pengolahan data teks yang terkait
CPMK-2	Mampu melakukan penggalian data teks dari sumber digital dan mengolahnya menggunakan teknik pre-processing, ekstraksi fitur, dan similarity teks
CPMK-3	Mampu melakukan pemodelan data teks dengan menggunakan klasifikasi dan klaster
CPMK-4	Mampu merepresentasikan pengetahuan linguistik pada tingkat representasi morfologi, sintaksis serta semantik

Matrik CPL - CPMK

CPMK	CPL-3	CPL-19	CPL-22	CPL-26
CPMK-1	✔			
CPMK-2		✔		
CPMK-3			✔	
CPMK-4				✔

Matrik CPMK pada Kemampuan akhir tiap tahapan belajar (Sub-CPMK)

CPMK	Minggu Ke															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
CPMK-1												✔	✔	✔	✔	✔
CPMK-2		✔	✔	✔	✔	✔	✔	✔								
CPMK-3									✔	✔	✔					
CPMK-4	✔															

Pustaka

Utama :

1. Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019

Pendukung :

1. Daniel Jurafsky & James H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2nd Edition, Prentice Hall, 2008.

Mg Ke-	Kemampuan akhir tiap tahapan belajar (Sub-CPMK)	Penilaian		Bentuk Pembelajaran, Metode Pembelajaran, Penugasan Mahasiswa, [Estimasi Waktu]		Materi Pembelajaran	Bobot Penilaian (%)
		Indikator	Kriteria & Bentuk	Luring (*offline*)	*Daring (online)*		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	Memahami konsep morfologi bahasa mulai dari fonem, morpheme, lexemes, syntax, dan contex	1. Menjelaskan konsep morfologi bahasa manusia 2. Menjelaskan konsep phoneme 3. Menjelaskan konsep morpheme dan lexeme 4. Menjelaskan konsep syntax 5. Menjelaskan konsep context	Non-Tes Tugas	Pembelajaran Kolaboratif (Ceramah, diskusi, dan tanya jawab)			2%
2	Menjelaskan represesntasi string, operasi string, regex, dan visualisasi data teks pada pemrograman	1. Menjelaskan representasi string. 2. Menjelaskan operasi pada string 3. Menjelaskan teknik indexing dan slicing 4. Menjelaskan method pada string 5. Menjelaskan formatting pada string 6. Menjelaskan regex pada string	Non-Tes Tugas	Pembelajaran Kolaboratif (Ceramah, diskusi, dan tanya jawab)			2%
3	Melakukan Web Scraping (HTTP dan API)	1. Mahasiswa mampu menjelaskan definisi dari web scraping 2. Mahasiswa mampu melakukan web scraping dengan HTTP 3. Mahasiswa mampu melakukan web scraping dengan API	Non-Tes Tugas	Pembelajaran Kolaboratif (Ceramah, diskusi, dan tanya jawab)			2%
4	Mahasiswa mampu melakukan teknik pre-processing pada data teks	1. Mahasiswa mampu melakukan teknik preprocessing data teks sederhana (tag HTML, stopwords, dll) 2. Mahasiswa mampu melakukan teknik preprocessing advance (parsing, casefolding, stemming, dll)	Non-Tes Tugas	Pembelajaran Kolaboratif (Ceramah, diskusi, tanya jawab)			2%

5 dan 6	Melakukan teknik ekstraksi fitur pada data teks	1. Mahasiswa mampu melakukan ekstraksi fitur dengan One Hot Encoding 2. Mahasiswa mampu melakukan ekstraksi fitur dengan Bag of Words 3. Mahasiswa mampu melakukan ekstraksi fitur dengan TF-IDF 4. Mahasiswa mampu melakukan ekstraksi fitur dengan Word2Vec 5. Mahasiswa mampu melakukan ekstraksi fitur dengan Fast-Text 6. Mahasiswa mampu melakukan ekstraksi fitur dengan Glove	Tugas	Pembelajaran Kolaboratif (Ceramah, diskusi, dan tanya jawab)			2.5%
7	Mahasiswa dapat melakukan teknik dalam similaritas dari data teks	1. Mahasiswa mampu melakukan teknik similarity text dengan term dan document 2. Mahasiswa mampu melakukan teknik similarity text dengan konsep jarak 3. Mahasiswa mampu melakukan teknik similarity text untuk recommender system	Non-Tes	Pembelajaran Kolaboratif (Ceramah, diskusi, dan tanya jawab)			3%
8	Mahasiswa mampu mengimplementasikan pada mini case yang terjadi di kehidupan sehari-hari	Ujian Tengah Semester	Non-Tes	Ujian Tengah Semester (UTS)			5%
9	Mahasiswa mampu melakukan teknik clustering text pada data teks	1. Mahasiswa mampu melakukan text clustering dengan hierarchical clustering 2. Mahasiswa mampu melakukan text clustering dengan non hierarchical clustering 3. Mahasiswa mampu melakukan text clustering dengan density clustering model	Non-Tes	Pembelajaran Kolaboratif (Ceramah, diskusi, dan tanya jawab)			3%
10	Mahasiswa mampu mendapatkan topik utama dengan LDA	1. Mahasiswa mampu mengaplikasikan LDA pada data text 2. Mahasiswa mampu menjelaskan hasil LDA dari data teks	Non-Tes	Pendekatan kolaboratif (diskusi dan ekspositori) Diskusi			3%
11	Mahasiswa mampu menerapkan konsep NLP (pada summarization)	1. Mahasiswa mampu menjelaskan konsep NLP 2. Mahasiswa mampu melakukan summarization pada data teks	Non-Tes	Pendekatan kolaboratif (diskusi dan ekspositori) Diskusi			3%
12 s.d 15	Mampu merancang penyelesaian masalah pada data teks menggunakan pengolahan data teks yang terkait	1. Menentukan topik permasalahan 2. Merancang ide penyelesaian 3. Presentasi Laporan Kemajuan	Presentasi Proposal dan Laporan Kemajuan	Aktifitas Partisipatif, Praktik / Unjuk Kerja			10%

16	Mampu mengimplementasikan penyelesaian masalah pada data teks menggunakan pengolahan data teks yang terkait	1. Mempresentasikan hasil project 2. Kepemahaman teori dari hasil project	Penilaian Hasil Akhir Project	Aktifitas Partisipatif, Praktik / Unjuk Kerja			30%

Catatan

1. **Capaian Pembelajaran Lulusan Program Studi (CPL-Prodi)** adalah rumusan kemampuan yang harus dimiliki oleh lulusan program studi yang mencakup sikap, penguasaan pengetahuan, keterampilan umum, dan keterampilan khusus sesuai dengan jenjang kualifikasi, yang diperoleh melalui proses pembelajaran.
2. **CPL-Prodi yang dibebankan pada Mata Kuliah** adalah bagian dari CPL-Prodi yang dipilih dan relevan untuk dikembangkan melalui suatu mata kuliah, sebagai dasar perumusan capaian pembelajaran mata kuliah.
3. **Capaian Pembelajaran Mata Kuliah (CPMK)** adalah kemampuan yang dirumuskan secara spesifik dan diturunkan dari CPL-Prodi yang dibebankan pada mata kuliah, serta menggambarkan kompetensi yang harus dicapai mahasiswa setelah menyelesaikan mata kuliah.
4. **Sub-CPMK** adalah kemampuan akhir yang direncanakan pada setiap tahap pembelajaran, sebagai penjabaran lebih rinci dari CPMK, bersifat terukur dan dapat diamati sesuai dengan materi pembelajaran.
5. **Indikator Penilaian** adalah pernyataan spesifik dan terukur yang digunakan untuk menilai ketercapaian Sub-CPMK, baik dari proses maupun hasil belajar mahasiswa, yang didukung dengan bukti penilaian.
6. **Kriteria Penilaian** adalah tolok ukur yang digunakan untuk menentukan tingkat ketercapaian hasil belajar mahasiswa berdasarkan indikator penilaian, guna menjamin penilaian yang objektif, konsisten, dan adil. Kriteria penilaian dapat bersifat kuantitatif dan/atau kualitatif.
7. **Bentuk Penilaian** meliputi penilaian tes dan non-tes sesuai dengan karakteristik capaian pembelajaran.
8. **Bentuk Pembelajaran** meliputi kuliah, responsi, tutorial, seminar atau bentuk lain yang setara, praktikum, praktik studio, praktik bengkel, praktik lapangan, penelitian, dan pengabdian kepada masyarakat.
9. **Metode Pembelajaran** merupakan cara penyampaian pembelajaran yang digunakan untuk mencapai capaian pembelajaran, antara lain Small Group Discussion, Project Based Learning, Case Based Learning, Collaborative Learning, Discovery Learning, Self-Directed Learning, serta metode lain yang relevan.
10. **Materi Pembelajaran** adalah bahan kajian yang disusun secara sistematis dalam bentuk pokok dan sub-pokok bahasan untuk mendukung pencapaian Sub-CPMK.
11. **Bobot Penilaian** adalah persentase kontribusi setiap komponen penilaian terhadap nilai akhir mata kuliah, yang ditetapkan secara proporsional sesuai tingkat kompleksitas dan ketercapaian Sub-CPMK, dengan total keseluruhan sebesar 100%.
12. TM = Tatap Muka, PT = Penugasan Terstruktur, BM = Belajar Mandiri.