



**Universitas Negeri Surabaya**  
**Fakultas Matematika dan Ilmu Pengetahuan Alam**  
**Program Studi S1 Sains Data**

Kode Dokumen

**RENCANA PEMBELAJARAN SEMESTER**

<b>MATA KULIAH (MK)</b>	<b>KODE</b>	<b>Rumpun MK</b>	<b>BOBOT (sks)</b>	<b>SEMESTER</b>	<b>Tgl Penyusunan</b>																																																																																																					
Pemrosesan Teks	4920202030		T=3 P=0 ECTS=4.77	3	29 Juli 2023																																																																																																					
<b>OTORISASI</b>	<b>Pengembang RPS</b>		<b>Koordinator RMK</b>		<b>Koordinator Program Studi</b>																																																																																																					
	Riskyana Dewi Intan Puspitasari, M.Kom		Dr. Elly Matul Imah., M.Kom		Yuliani Puji Astuti, S.Si., M.Si.																																																																																																					
<b>Model Pembelajaran</b>	Project Based Learning																																																																																																									
<b>Capaian Pembelajaran (CP)</b>	<b>CPL-PRODI yang dibebankan pada MK</b>																																																																																																									
	<b>CPL-9</b>	Mampu menerapkan prinsip-prinsip sains data untuk menyelesaikan masalah																																																																																																								
	<b>CPL-16</b>	Menguasai teori dan konsep sains data																																																																																																								
	<b>Capaian Pembelajaran Mata Kuliah (CPMK)</b>																																																																																																									
	<b>CPMK - 1</b>	Menjelaskan konsep Language Modeling dalam pengolahan teks																																																																																																								
	<b>CPMK - 2</b>	Mampu merepresentasikan pengetahuan linguistik pada tingkat representasi morfologi, sintaksis serta semantik																																																																																																								
	<b>CPMK - 3</b>	Mampu melakukan penggalian data teks dari sumber digital dan mengolahnya menggunakan teknik pre-processing, ekstraksi fitur, dan klasifikasi teks																																																																																																								
	<b>CPMK - 4</b>	Mampu merancang penyelesaian masalah pada data teks menggunakan pengolahan data teks yang terkait																																																																																																								
	<b>Matrik CPL - CPMK</b>																																																																																																									
		<table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>CPMK</th> <th>CPL-9</th> <th>CPL-16</th> </tr> </thead> <tbody> <tr><td>CPMK-1</td><td></td><td></td></tr> <tr><td>CPMK-2</td><td></td><td></td></tr> <tr><td>CPMK-3</td><td></td><td></td></tr> <tr><td>CPMK-4</td><td></td><td></td></tr> </tbody> </table>					CPMK	CPL-9	CPL-16	CPMK-1			CPMK-2			CPMK-3			CPMK-4																																																																																							
CPMK	CPL-9	CPL-16																																																																																																								
CPMK-1																																																																																																										
CPMK-2																																																																																																										
CPMK-3																																																																																																										
CPMK-4																																																																																																										
<b>Matrik CPMK pada Kemampuan akhir tiap tahapan belajar (Sub-CPMK)</b>																																																																																																										
	<table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th rowspan="2">CPMK</th> <th colspan="16">Minggu Ke</th> </tr> <tr> <th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th><th>9</th><th>10</th><th>11</th><th>12</th><th>13</th><th>14</th><th>15</th><th>16</th> </tr> </thead> <tbody> <tr><td>CPMK-1</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>CPMK-2</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>CPMK-3</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> <tr><td>CPMK-4</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr> </tbody> </table>					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																
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																																																																																																										
<b>Deskripsi Singkat MK</b>	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 seperti similarity dan klasterisasi teks. Topik-topik yang dibahas meliputi: morfologi bahasa, representasi 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.																																																																																																									
<b>Pustaka</b>	<b>Utama :</b>	1. Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019																																																																																																								
	<b>Pendukung :</b>	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.																																																																																																								
<b>Dosen Pengampu</b>	Dr. Elly Matul Imah, M.Kom. Riskyana Dewi Intan Puspitasari, M.Kom.																																																																																																									
	<b>Penilaian</b>	<b>Bantuk Pembelajaran, Metode Pembelajaran, Penugasan Mahasiswa, [ Estimasi Waktu]</b>																																																																																																								

Mg Ke-	Kemampuan akhir tiap tahapan belajar (Sub-CPMK)					Materi Pembelajaran [ Pustaka ]	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 context	<ol style="list-style-type: none"> <li>1. Menjelaskan konsep morfologi bahasa manusia</li> <li>2. Menjelaskan konsep phoneme</li> <li>3. Menjelaskan konsep morpheme dan lexeme</li> <li>4. Menjelaskan konsep syntax</li> <li>5. Menjelaskan konsep context</li> </ol>	<p><b>Kriteria:</b> Non-Tes Tugas</p> <p><b>Bentuk Penilaian :</b> Aktifitas Partisipatif</p>	Pembelajaran Kolaboratif (Ceramah, diskusi, dan tanya jawab) 3 X 50		<p><b>Materi:</b> Konsep Bahasa, Phoneme, Morpheme dan Lexeme, Syntax, Context</p> <p><b>Pustaka:</b> <i>Daniel Jurafsky &amp; James H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 2nd Edition, Prentice Hall, 2008.</i></p>	2%
2	Menjelaskan representasi string, operasi string, regex, dan visualisasi data teks pada pemrograman	<ol style="list-style-type: none"> <li>1. Menjelaskan representasi string.</li> <li>2. Menjelaskan operasi pada string</li> <li>3. Menjelaskan teknik indexing dan slicing</li> <li>4. Menjelaskan method pada string</li> <li>5. Menjelaskan formatting pada string</li> <li>6. Menjelaskan regex pada string</li> </ol>	<p><b>Kriteria:</b> Non-Tes Tugas</p> <p><b>Bentuk Penilaian :</b> Aktifitas Partisipatif</p>	Pembelajaran Kolaboratif (Ceramah, diskusi, dan tanya jawab) 3 X 50		<p><b>Materi:</b> Representasi string, Operasi string, Indexing dan slicing, String method, String formatting, String regex</p> <p><b>Pustaka:</b> <i>Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019</i></p>	2%
3	Melakukan teknik pre-prosesing data teks	<ol style="list-style-type: none"> <li>1. Menghapus tag HTML</li> <li>2. Melakukan tokenisasi</li> <li>3. Menghapus stopwords</li> <li>4. Mengatasi spelling error pada teks</li> </ol>	<p><b>Kriteria:</b> Non-Tes Tugas</p> <p><b>Bentuk Penilaian :</b> Aktifitas Partisipatif</p>	Pembelajaran Kolaboratif (Ceramah, diskusi, dan tanya jawab) 3 X 50		<p><b>Materi:</b> Menghapus tag HTML, Melakukan tokenisasi, Menghapus stopwords, Mengatasi spelling error pada teks</p> <p><b>Pustaka:</b> <i>Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019</i></p>	2%

4	Melakukan teknik pre-prosesing data teks	<ol style="list-style-type: none"> <li>1.Melakukan Stemming.</li> <li>2.Melakukan Lemmatization</li> <li>3.Melakukan Tagging</li> <li>4. Melakukan Chunking</li> <li>5.Melakukan Parsing</li> </ol>	<b>Kriteria:</b> Tugas  <b>Bentuk Penilaian :</b> Aktifitas Partisipasif / Unjuk Kerja	Pembelajaran Kolaboratif (Ceramah, diskusi, dan tanya jawab) 3 X 50		<b>Materi:</b> Stemming, Lemmatization, Tagging, Chunking, Parsing <b>Pustaka:</b> Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019	10%
5	Melakukan teknik ekstraksi fitur pada data teks	<ol style="list-style-type: none"> <li>1.Melakukan teknik Bag of Words</li> <li>2.Melakukan teknik Bag of N-grams</li> <li>3.Melakukan Parsing</li> </ol>	<b>Kriteria:</b> Tugas  <b>Bentuk Penilaian :</b> Aktifitas Partisipasif	Pembelajaran Kolaboratif (Ceramah, diskusi, dan tanya jawab) 3 X 50		<b>Materi:</b> Bag of Words, Bag of N-gram Words <b>Pustaka:</b> Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019	2%
6	Melakukan teknik ekstraksi fitur pada data teks	<ol style="list-style-type: none"> <li>1.Melakukan teknik TF-IDF</li> <li>2. Melakukan teknik similarity features</li> </ol>	<b>Kriteria:</b> Non-Tes  <b>Bentuk Penilaian :</b> Aktifitas Partisipasif	Pembelajaran Kolaboratif (Ceramah, diskusi, dan tanya jawab) 3 X 50		<b>Materi:</b> TF-IDF, Similarity Features <b>Pustaka:</b> Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019	2%
7	Melakukan teknik ekstraksi fitur pada data teks	<ol style="list-style-type: none"> <li>1.Melakukan teknik Word2Vec</li> <li>2.Melakukan teknik Glove</li> <li>3.Melakukan teknik FastText</li> </ol>	<b>Kriteria:</b> Non-Tes  <b>Bentuk Penilaian :</b> Aktifitas Partisipasif	Pembelajaran Kolaboratif (Ceramah, diskusi, dan tanya jawab) 3 X 50		<b>Materi:</b> Word2Vec, Glove, FastText <b>Pustaka:</b> Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019	2%
8		Mampu menjawab pertanyaan dengan tepat dan benar	<b>Kriteria:</b> Ujian Tes Tulis  <b>Bentuk Penilaian :</b> Praktik / Unjuk Kerja, Tes	Ujian Tengah Semester (UTS) 100 menit		<b>Materi:</b> Semua materi sebelum UTS <b>Pustaka:</b> Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019	10%

9	Melakukan teks similarity pada data teks	<ol style="list-style-type: none"> <li>Melakukan teks similarity antar term dan document</li> <li>Menemukan dokumen paling relevan menggunakan teknik jarak</li> <li>Membuat rekomen sistem dari teknik teks similarity</li> </ol>	<b>Kriteria:</b> Non-Tes	Pendekatan kolaboratif (diskusi dan ekspositori) Diskusi 3x50		<b>Materi:</b> Character vectorization, Hamming Distance, Manhattan Distance, Euclidean Distance, Cosine Distance, Document Similarity, Recomen System <b>Pustaka:</b> Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019	2%
10	Melakukan teks clustering pada data teks	<ol style="list-style-type: none"> <li>Melakukan text clustering menggunakan hierarchical clustering model</li> <li>Melakukan text clustering menggunakan distribution clustering model</li> <li>Melakukan text clustering menggunakan density clustering model</li> </ol>	<b>Kriteria:</b> Non-Tes  <b>Bentuk Penilaian :</b> Aktifitas Partisipasif	Pendekatan kolaboratif (diskusi dan ekspositori) Diskusi 3x50		<b>Materi:</b> hierarchical clustering model, distribution clustering model, density clustering model <b>Pustaka:</b> Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019	2%
11	Melakukan Web Scraping (HTTP)	<ol style="list-style-type: none"> <li>Pengenalan library BeautifulSoup</li> <li>Mengenal metode dan atribut pada BeautifulSoup</li> <li>Melakukan crawling laman website</li> <li>Menyimpan hasil scraping data</li> </ol>	<b>Kriteria:</b> Non-Tes  <b>Bentuk Penilaian :</b> Aktifitas Partisipasif, Praktik / Unjuk Kerja	Pendekatan kolaboratif (diskusi dan ekspositori) Diskusi 3x50		<b>Materi:</b> Beautifulsoal, Website Layout, Crawling <b>Pustaka:</b> Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019	2%
12	Melakukan Web Scraping (API)	<ol style="list-style-type: none"> <li>Mengenal konsep API</li> <li>Melakukan parsing menggunakan JSON</li> <li>Melakukan crawling laman website melalui API</li> <li>Menyimpan hasil scraping data</li> </ol>	<b>Kriteria:</b> Tes  <b>Bentuk Penilaian :</b> Aktifitas Partisipasif, Praktik / Unjuk Kerja	Pendekatan kolaboratif (diskusi dan ekspositori) Diskusi 3x50		<b>Materi:</b> Scraping using HTTP, API, JSON <b>Pustaka:</b> Sarkar, Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019	10%

13	Mengimplementasikan teknik pemrosesan data pada permasalahan nyata	1. Menentukan topik permasalahan 2. Menyusun tim dan jadwal pengerjaan project 3. Membuat desain pemecahan masalah project	<b>Kriteria:</b> Presentasi  <b>Bentuk Penilaian :</b> Aktifitas Partisipasif, Penilaian Hasil Project / Penilaian Produk, Praktik / Unjuk Kerja	Aktifitas Partisipasif, Praktik / Unjuk Kerja 3x50		<b>Materi:</b> Semua materi <b>Pustaka:</b> Sarkar, <i>Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019</i>	10%
14	Mengimplementasikan teknik pemrosesan data pada permasalahan nyata	1. Mengimplementasikan hasil project 2. Merealisasikan hasil project yang telah disetujui	<b>Kriteria:</b> Presentasi  <b>Bentuk Penilaian :</b> Aktifitas Partisipasif, Penilaian Hasil Project / Penilaian Produk	Aktifitas Partisipasif, Praktik / Unjuk Kerja 3x50		<b>Materi:</b> Semua materi <b>Pustaka:</b> Sarkar, <i>Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019</i>	5%
15	Mengimplementasikan teknik pemrosesan data pada permasalahan nyata	Presentasi kemajuan dan laporan pencapaian kerja kelompok	<b>Kriteria:</b> Presentasi  <b>Bentuk Penilaian :</b> Aktifitas Partisipasif, Penilaian Hasil Project / Penilaian Produk	Aktifitas Partisipasif, Praktik / Unjuk Kerja 3x50		<b>Materi:</b> Semua materi <b>Pustaka:</b> Sarkar, <i>Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019</i>	9%
16	Ujian Akhir Semester (UAS)	Presentasi dan tanya-jawab	<b>Kriteria:</b> Aktifitas Partisipasif, Penilaian Hasil Project / Penilaian Produk, Penilaian Portofolio  <b>Bentuk Penilaian :</b> Aktifitas Partisipasif, Penilaian Hasil Project / Penilaian Produk, Penilaian Portofolio	Presentasi dan tanya jawab 3x50		<b>Materi:</b> Semua materi <b>Pustaka:</b> Sarkar, <i>Dipanjan, Text Analytics with Python A Practitioner's Guide to Natural Language Processing, Second Edition, Apress, 2019</i>	30%

#### Rekap Persentase Evaluasi : Project Based Learning

No	Evaluasi	Persentase
1.	Aktifitas Partisipasif	45.33%
2.	Penilaian Hasil Project / Penilaian Produk	20.33%
3.	Penilaian Portofolio	10%
4.	Praktik / Unjuk Kerja	19.33%
5.	Tes	5%
		99.99%

#### Catatan

1. **Capaian Pembelajaran Lulusan PRODI (CPL-PRODI)** adalah kemampuan yang dimiliki oleh setiap lulusan PRODI yang merupakan internalisasi dari sikap, penguasaan pengetahuan dan ketrampilan sesuai dengan jenjang prodinya yang diperoleh melalui proses pembelajaran.
2. **CPL yang dibebankan pada mata kuliah** adalah beberapa capaian pembelajaran lulusan program studi (CPL-PRODI) yang digunakan untuk pembentukan/pengembangan sebuah mata kuliah yang terdiri dari aspek sikap, ketrampilan umum, ketrampilan khusus dan pengetahuan.
3. **CP Mata kuliah (CPMK)** adalah kemampuan yang dijabarkan secara spesifik dari CPL yang dibebankan pada mata kuliah, dan bersifat spesifik terhadap bahan kajian atau materi pembelajaran mata kuliah tersebut.
4. **Sub-CP Mata kuliah (Sub-CPMK)** adalah kemampuan yang dijabarkan secara spesifik dari CPMK yang dapat diukur atau diamati dan merupakan kemampuan akhir yang direncanakan pada tiap tahap pembelajaran, dan bersifat spesifik terhadap materi pembelajaran mata kuliah tersebut.
5. **Indikator penilaian** kemampuan dalam proses maupun hasil belajar mahasiswa adalah pernyataan spesifik dan terukur yang mengidentifikasi kemampuan atau kinerja hasil belajar mahasiswa yang disertai bukti-bukti.
6. **Kreteria Penilaian** adalah patokan yang digunakan sebagai ukuran atau tolok ukur ketercapaian pembelajaran dalam penilaian berdasarkan indikator-indikator yang telah ditetapkan. Kreteria penilaian merupakan pedoman bagi penilai agar penilaian konsisten

- dan tidak bias. Kreteria dapat berupa kuantitatif ataupun kualitatif.
7. **Bentuk penilaian:** tes dan non-tes.
  8. **Bentuk pembelajaran:** Kuliah, Responsi, Tutorial, Seminar atau yang setara, Praktikum, Praktik Studio, Praktik Bengkel, Praktik Lapangan, Penelitian, Pengabdian Kepada Masyarakat dan/atau bentuk pembelajaran lain yang setara.
  9. **Metode Pembelajaran:** Small Group Discussion, Role-Play & Simulation, Discovery Learning, Self-Directed Learning, Cooperative Learning, Collaborative Learning, Contextual Learning, Project Based Learning, dan metode lainnya yg setara.
  10. **Materi Pembelajaran** adalah rincian atau uraian dari bahan kajian yg dapat disajikan dalam bentuk beberapa pokok dan sub-pokok bahasan.
  11. **Bobot penilaian** adalah prosentasi penilaian terhadap setiap pencapaian sub-CPMK yang besarnya proposional dengan tingkat kesulitan pencapaian sub-CPMK tsb., dan totalnya 100%.
  12. TM=Tatap Muka, PT=Penugasan terstruktur, BM=Belajar mandiri.

RPS ini telah divalidasi pada tanggal 3 Mei 2024

Koordinator Program Studi S1  
Sains Data



Yuliani Puji Astuti, S.Si., M.Si.  
NIDN 0031077804

UPM Program Studi S1 Sains  
Data



Riskyana Dewi Intan Puspitasari,  
M.Kom.  
NIDN 0021059403

File PDF ini digenerate pada tanggal 5 Juli 2024 Jam 00:30 menggunakan aplikasi RPS-OBE SiDia Unesa

**VALID**