



**INSTITUT TEKNOLOGI SEPULUH NOPEMBER, FAKULTAS SAINS DAN
ANALITIKA DATA, PRODI SAINS DATA STATISTIKA**

**Kode
Dokumen**

RENCANA PEMBELAJARAN SEMESTER

MATA KULIAH (MK)	KODE	Rumpun MK	BOBOT (sks)		SEMESTER	Tgl Penyusunan
Komputasi Sains Data	SD234402	Statistika Komputasi dan Sains Data	T=2	P=1	4	
OTORISASI	Pengembang RPS		Koordinator RMK		Ketua PRODI	
	Veniola Forestryani, S.Si., M.Si.; Erma Oktania Permatasari, S.Si., M.Si.; Widhianingsih Tintrim Dwi Ary, S.Si., M.Stat., Ph.D. Prof. Drs. Nur Iriawan, M.Ikom., Ph.D.		Prof. Drs. Nur Iriawan, M.Ikom., Ph.D.		Dr. Achmad Choiruddin, S.Si., M.Sc.	
Capaian Pembelajaran (CP)	CPL-PRODI yang dibebankan pada MK					
	CPL-5	Mampu menerapkan teori dan metode statistika pada analisis data untuk mendukung pengambilan keputusan				
	CPL-6	Mampu merancang pengumpulan dan pengintegrasian data terstruktur dan tidak terstruktur dengan metodologi yang tepat				
	CPL-7	Mampu menyusun algoritma dan membuat program komputer untuk pengolahan data berskala besar dalam penyelesaian masalah di berbagai bidang terapan				
	Capaian Pembelajaran Mata Kuliah (CPMK)					
	CPMK-1	Mampu menjelaskan jenis-jenis data terstruktur dan tidak terstruktur serta perbedaan di antara keduanya, serta melakukan manipulasi dan pembersihan data tersebut menggunakan R dan Python.				
	CPMK-2	Mampu menjelaskan dan menerapkan metode visualisasi data terstruktur menggunakan <i>library</i> visualisasi di R dan Python untuk membuat visualisasi data yang efektif.				
	CPMK-3	Mampu memahami dan mengidentifikasi distribusi data terstruktur menggunakan alat statistik di R dan Python, serta mengimplementasikan analisis distribusi dalam sains data.				

	CPMK-4	Mampu menjelaskan teknik-teknik pengolahan data tidak terstruktur, termasuk analisis teks, menggunakan R dan Python, dan mampu mengimplementasikan analisis tersebut dalam sains data.																																
	CPMK-5	Mampu menjelaskan dan mengimplementasikan algoritma <i>machine learning</i> untuk data terstruktur menggunakan scikit-learn di Python dan caret di R, termasuk klasifikasi, regresi, dan clustering.																																
	CPMK-6	Mampu menjelaskan dasar-dasar analisis big data dan menggunakan alat seperti Hadoop dan Spark untuk analisis data tidak terstruktur dengan PySpark di Python dan sparklyr di R.																																
	CPMK-7	Mampu mengkomunikasikan hasil analisis data secara efektif dan bekerja sama dalam tim interdisipliner dan multidisipliner.																																
		<p>Matrik CPL – CPMK</p> <table border="1"> <thead> <tr> <th>CPMK</th> <th>CPL-5</th> <th>CPL-6</th> <th>CPL-7</th> </tr> </thead> <tbody> <tr> <td>CPMK-1</td> <td></td> <td>V</td> <td></td> </tr> <tr> <td>CPMK-2</td> <td>V</td> <td>V</td> <td>V</td> </tr> <tr> <td>CPMK-3</td> <td>V</td> <td>V</td> <td>V</td> </tr> <tr> <td>CPMK-4</td> <td>V</td> <td>V</td> <td>V</td> </tr> <tr> <td>CPMK-5</td> <td>V</td> <td>V</td> <td>V</td> </tr> <tr> <td>CPMK-6</td> <td>V</td> <td>V</td> <td>V</td> </tr> <tr> <td>CPMK-7</td> <td></td> <td></td> <td>V</td> </tr> </tbody> </table>	CPMK	CPL-5	CPL-6	CPL-7	CPMK-1		V		CPMK-2	V	V	V	CPMK-3	V	V	V	CPMK-4	V	V	V	CPMK-5	V	V	V	CPMK-6	V	V	V	CPMK-7			V
CPMK	CPL-5	CPL-6	CPL-7																															
CPMK-1		V																																
CPMK-2	V	V	V																															
CPMK-3	V	V	V																															
CPMK-4	V	V	V																															
CPMK-5	V	V	V																															
CPMK-6	V	V	V																															
CPMK-7			V																															
Deskripsi Singkat MK	Mata kuliah ini bertujuan untuk membekali mahasiswa dengan kemampuan pemrograman R dan Python kemudian memformulasikan penyelesaian masalah menggunakan dasar-dasar algoritma komputasi terkait dengan Sistem Komputasi, Analisis , dan Sains Data. Mata kuliah ini menitik-beratkan pada peranan komputasi sebagai alat fundamental dalam algoritma analitik, analisis interaktif (Jupyter Notebook, dll), dan teknik visualisasi untuk analisis data. Materi disampaikan melalui ceramah interaktif, diskusi, latihan, praktikum serta <i>Problem Based Learning</i> .																																	
Bahan Kajian: Materi Pembelajaran	BK-5 Teori dan Metode Statistika BK-6 Pengumpulan dan Manajemen Data BK-7 Program Komputer dan Analisis Big Data																																	
Pustaka	Utama :																																	
	<ol style="list-style-type: none"> Bruce, Peter, Andrew Bruce, and Peter Gedeck. Practical statistics for data scientists: 50+ essential concepts using R and Python. O'Reilly Media, 2020. Piegorsch, Walter W., et al., eds. Computational Statistics in Data Science. John Wiley & Sons, 2022. VanderPlas, Jake, Python Data Science Handbook: Essential Tools for Working with Data, O'Reilly Media, 2016. 																																	

		Pendukung :					
				<ol style="list-style-type: none"> 1. Grus, Joel, Data Science from Scratch: First Principles with Python, O'Reilly Media, 2015. 2. Hill, Christian, Learning Scientific Programming with Python, Cambridge University Press, 2016. 3. McKinney, Wes, Python for Data Analysis: Data Wrangling with Pandas, NumPy and IPython. 2nd Edition, O'Reilly Media, 2017. 4. NumPy User Guide (https://docs.scipy.org/doc/numpy/user/index.html) 5. Ramalho, Luciano, Fluent Python: Clear, Concise, and Effective Programming, O'Reilly Media, 2015. 6. Severance, Charles, Python for Informatics: Exploring Information, CreateSpace Independent Publishing Platform, 2013. 7. SciPy Lecture Notes (http://www.scipy-lectures.org/) 			
Dosen Pengampu		Veniola Forestryani, S.Si., M.Si.; Erma Oktania Permatasari, S.Si., M.Si.; Widhianingsih Tintrim Dwi Ary, S.Si., M.Stat., Ph.D.					
Matakuliah syarat		Pemrograman Open Source					
Mg Ke-	Kemampuan akhir tiap tahapan belajar (Sub-CPMK)	Penilaian		Bantuk Pembelajaran, Metode Pembelajaran, Penugasan Mahasiswa, [Estimasi Waktu]		Materi Pembelajaran [Pustaka]	Bobot Penilaian (%)
		Indikator	Kriteria & Bentuk	Luring (<i>offline</i>)	Daring (<i>online</i>)		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	Pengenalan Sains Data dan Jenis Data	Mampu mengidentifikasi jenis-jenis data yang biasa digunakan dalam sains data, antara lain <i>structured (Numeric & Categorical)</i> dan <i>unstructured</i> (gambar, teks, <i>Clickstreams</i> , dll)	Tugas, Observasi di kelas, Tes, Presentasi Makalah	Ceramah Interaktif, Diskusi, Praktikum TM: 1×[3x50"] BM: 1×[3×60"] PT: 1×[3×60"]		<i>Exploratory Data Analysis</i> - Elemen Data Terstruktur - Elemen Data Tidak Terstruktur Bruce, Peter, Andrew Bruce, and Peter	5%

						Gedeck. <i>Practical statistics for data scientists: 50+ essential concepts using R and Python</i> . O'Reilly Media, 2020. (Ch. 1)	
2	Mampu melakukan manipulasi dan pembersihan data terstruktur menggunakan teknik yang tepat di R dan Python	<ul style="list-style-type: none"> - Mampu mengidentifikasi data yang perlu dibersihkan dan disiapkan untuk analisis - Menggunakan <i>library</i> Pandas di Python atau dplyr dan tidyr di R untuk manipulasi dan pembersihan data - Mengimplementasikan teknik pembersihan data seperti handling missing values, data normalization, dan data transformation 	Tugas, Observasi di kelas, Tes, Presentasi Makalah	Ceramah Interaktif, Diskusi, Praktikum TM: 1×[2x50"] P: 1×[1x170"] BM: 1×[2×60"] PT: 1×[2×60"]		<ul style="list-style-type: none"> - Teknik Pembersihan Data - Teknik Persiapan Data - Penggunaan Pandas (Python) - Penggunaan dplyr dan tidyr (R) <p>McKinney, Wes. Python for Data Analysis. O'Reilly Media, 2017. (Ch. 2)</p>	10%
3-4	<ul style="list-style-type: none"> - Mampu membuat visualisasi data terstruktur yang efektif menggunakan Matplotlib/Seaborn di Python atau ggplot2 di R - Mampu mengidentifikasi distribusi data terstruktur menggunakan alat statistik di R dan Python 	<ul style="list-style-type: none"> - Mampu menggunakan pustaka visualisasi seperti Matplotlib dan Seaborn di Python atau ggplot2 di R untuk membuat visualisasi yang informatif - Mampu mengidentifikasi dan memvisualisasikan distribusi data menggunakan SciPy di Python atau fitdistrplus di R 	Tugas, Observasi di kelas, Tes, Presentasi Makalah	Ceramah Interaktif, Diskusi, Praktikum TM: 2×[2x50"] P: 2×[1x170"] BM: 2×[2×60"] PT: 2×[2×60"]		<ul style="list-style-type: none"> - Teknik Visualisasi Data - Penggunaan Matplotlib dan Seaborn (Python) - Penggunaan ggplot2 (R) - Identifikasi Distribusi Data - Penggunaan SciPy (Python) - Penggunaan fitdistrplus (R) <p>McKinney, Wes. Python for Data Analysis.</p>	15%

		- Mampu mengimplementasikan teknik visualisasi dan analisis distribusi dalam proyek data science				O'Reilly Media, 2017. (Ch. 9-11) Bruce, Peter, Andrew Bruce, and Peter Gedeck. Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python. O'Reilly Media, 2020. (Ch. 2)	
5-6	Mampu melakukan pengolahan data tidak terstruktur termasuk analisis teks dan sentimen menggunakan R dan Python	- Mampu menggunakan <i>library</i> Pandas di Python atau tidytext di R untuk pengolahan data tidak terstruktur - Mampu menggunakan NLTK di Python atau syuzhet dan tm di R untuk analisis teks dan sentimen - Mampu mengimplementasikan teknik pengolahan dan analisis data tidak terstruktur dalam proyek data science	Tugas, Observasi di kelas, Tes, Presentasi Makalah	Ceramah Interaktif, Diskusi, Praktikum TM: 2x[2x50"] P: 2x[1x170"] BM: 2x[2x60"] PT: 2x[2x60"]		- Teknik Pengolahan Data Tidak Terstruktur - Analisis Teks dan Sentimen - Penggunaan Pandas (Python) - Penggunaan tidytext (R) - Penggunaan NLTK (Python) - Penggunaan syuzhet dan tm (R) Silge, Julia, and David Robinson. Text Mining with R: A Tidy Approach. O'Reilly Media, 2017. (Ch. 1-4) Bird, Steven, Ewan Klein, and Edward Loper. Natural Language Processing with Python. O'Reilly Media, 2009. (Ch. 1-6)	10%
7	Mampu menjelaskan dan mengimplementasikan algoritma <i>machine learning</i>	- Mampu menggunakan <i>library</i> scikit-learn di Python atau caret di R	Tugas, Observasi di kelas,	Ceramah Interaktif, Diskusi, Praktikum		- Teknik Machine Learning - Algoritma Klasifikasi	10%

	untuk data terstruktur menggunakan R dan Python	<ul style="list-style-type: none"> untuk membangun model pembelajaran mesin - Mampu mengimplementasikan teknik pembelajaran mesin seperti klasifikasi dan regresi pada data terstruktur - Mampu mengevaluasi performa model menggunakan metrik yang sesuai 	Tes, Presentasi Makalah	TM: 1×[2x50"] P: 1×[1x170"] BM: 1×[2×60"] PT: 1×[2×60"]		<p>dan Regresi</p> <ul style="list-style-type: none"> - Penggunaan scikit-learn (Python) - Penggunaan caret (R) <p>Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, 2019. (Ch. 1-4)</p> <p>Kuhn, Max, and Kjell Johnson. Applied Predictive Modeling. Springer, 2013. (Ch. 6-8)</p>	
8	Evaluasi Tengah Semester / Ujian Tengah Semester						
9-10	Mampu menjelaskan dan mengimplementasikan metode clustering dan segmentasi untuk data terstruktur menggunakan R dan Python	<ul style="list-style-type: none"> - Mampu menggunakan <i>library</i> scikit-learn di Python atau caret di R untuk melakukan clustering dan segmentasi - Mampu mengimplementasikan teknik clustering seperti K-means dan hierarchical clustering pada data terstruktur - Mampu mengevaluasi hasil clustering menggunakan metrik yang sesuai 	Tugas, Observasi di kelas, Tes, Presentasi Makalah	<p>Ceramah Interaktif, Diskusi, Praktikum</p> TM: 1×[2x50"] P: 2×[1x170"] BM: 1×[2×60"] PT: 1×[2×60"]		<ul style="list-style-type: none"> - Teknik Clustering - K-means Clustering - Hierarchical Clustering - Penggunaan scikit-learn (Python) - Penggunaan caret (R) <p>James, Gareth, et al. An Introduction to Statistical Learning with Applications in R. Springer, 2013. (Ch. 10)</p> <p>Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O'Reilly Media, 2019. (Ch. 9-10)</p>	10%

11-13	Mampu menjelaskan dasar-dasar analisis big data dan menggunakan alat seperti Hadoop dan Spark untuk analisis data tidak terstruktur dengan PySpark di Python dan sparklyr di R	<ul style="list-style-type: none"> - Menggunakan Hadoop untuk menyimpan dan memproses data besar - Menggunakan Spark untuk analisis data tidak terstruktur - Mengimplementasikan teknik big data menggunakan PySpark di Python dan sparklyr di R - Mengevaluasi hasil analisis data besar menggunakan metrik yang sesuai 	Tugas, Observasi di kelas, Tes, Presentasi Makalah	<p>Ceramah Interaktif, Diskusi, Praktikum</p> <p>TM: 3×[2x50"] P: 3×[1x170"] BM: 3×[2×60"] PT: 3×[2×60"]</p>		<ul style="list-style-type: none"> - Teknik Analisis Big Data - Penggunaan Hadoop - Penggunaan Spark - Penggunaan PySpark (Python) - Penggunaan sparklyr (R) <p>Damji, Jules S., et al. Learning Spark: Lightning-Fast Data Analytics. O'Reilly Media, 2020. (Ch. 1-4)</p> <p>Karau, Holden, et al. Learning Spark. O'Reilly Media, 2015. (Ch. 1-3)</p>	15%
14-15	<ul style="list-style-type: none"> - Mampu menghubungkan dan mengintegrasikan data dari berbagai sumber menggunakan R dan Python - Mampu merancang dan mengerjakan studi kasus integrasi data terstruktur dan tidak terstruktur 	<ul style="list-style-type: none"> - Mampu menggunakan SQLAlchemy di Python atau DBI di R untuk koneksi database - Mampu mengintegrasikan data terstruktur dan tidak terstruktur dari berbagai sumber - Mampu merancang dan mengimplementasikan studi kasus integrasi data yang mencakup teknik pembersihan, persiapan, dan analisis data 	Tugas, Observasi di kelas, Tes, Presentasi Makalah	<p>Ceramah Interaktif, Diskusi, Praktikum</p> <p>TM: 2×[2x50"] P: 2×[1x170"] BM: 2×[2×60"] PT: 2×[2×60"]</p>		<ul style="list-style-type: none"> - Menggunakan SQLAlchemy di Python atau DBI di R untuk koneksi database - Mengintegrasikan data terstruktur dan tidak terstruktur dari berbagai sumber - Merancang dan mengimplementasikan proyek integrasi data yang mencakup teknik pembersihan, persiapan, dan analisis data <p>McKinney, Wes. Python for Data Analysis.</p>	15%

						O'Reilly Media, 2017. (Ch. 6, 8, 9)	
						Wickham, Hadley, and Garrett Grolemund. R for Data Science. O'Reilly Media, 2016. (Ch. 10-12)	
16	Evaluasi Akhir Semester / Ujian Akhir Semester						

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. **Kriteria 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.