

# DATA WRANGLING

DSB 1224

## KONTRAK KULIAH

SENIN 7 MARET 2022

KTT1.8/ ONLINE

PROGRAM STUDI S1 SAINS DATA

FAKULTAS TEKNOLOGI INFORMASI DAN BISNIS

INSTITUT TEKNOLOGI TELKOM SURABAYA



## BIODATA DOSEN

**Nama** : Moh. Hamim Zajuli Al Faroby, S.Si., M.Mat.  
**NIDN/Kode** : 0724079501/HMM  
**E-mail** : alfaroby@ittelkom-sby.ac.id  
**Phone** : 081 331 653 603  
**Medsos** : @alfaroby\_mhz

### **PENELITIAN:**

2018

Identifikasi Jenis Kanker Darah (Leukemia) Terhadap Pengaruh Parameter Kernel Support Vector Machine dan Ekstraksi Ciri Rantai Markov Orde 2.

2019

Extreme Gradient Boosting Untuk Pencarian Protein yang Berpengaruh Terhadap Produksi Insulin Berdasarkan Interaksi Protein-Protein. [Penelitian Magister/ Hibah Penelitian DIKTI].

2020

Identifikasi Persebaran Avian Virus Influenza H9 di Indonesia Serta Prediksi Antiviralnya.  
[Penelitian Dasar/Hibah Penelitian DIKTI]

2021

Virtual Screening Senyawa Aktif Sebagai Inhibitor Meningitis Dengan Hybrid Molecular Fingerprint dan Pendekatan Machine Learning. [Penelitian Mandiri 2021]

**PUBLIKASI:**

2019

MI Irawan, MHZ Al Faroby, APD Nurhayati,” In Silico Analysis Using Hybrid Support Vector Machine and Second Order of Markov Chain for Multiple Sequence Alignment to Identify the Types of Leukaemia”, Journal of Physics: Conference Series 1366 (1), 012052, 2019.

2020

MHZ Al Faroby, MI Irawan, NNT Puspaningsih,” XGBoost and Network Analysis for Prediction of Proteins Affecting Insulin based on Protein Protein Interactions”, Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control (3), 1. 2020.

# PROFIL MATA KULIAH

Nama Matakuliah : Data Wrangling

Bobot : 4 SKS

Matakuliah *Data Wrangling* mempelajari dari tahapan awal bagaimana memproses data. Mata kuliah ini mengajarkan dasar operasi matriks dan vector menggunakan numpy dan proses pengolahan data terstruktur dalam bentuk data frame. Selain itu, matakuliah ini juga mempelajari bagaimana mendapatkan data dari sumber terbuka diinternet berupa web, secara otomatis. Serta dikenalkan jenis-jenis data yang terdapat di sumber terbuka seperti CSV, JSON dan XML. Setelah data terkumpul, matakuliah ini juga mengajarkan bagaimana mengeksplorasi data untuk baik itu membersihkan, mencari data hilang dan lain sebagainya. Selain itu, matakuliah ini juga mengajarkan bagaimana menormalisasi data dan menghasilkan data yang standar.

## **CAPAIAN PEMBELAJARAN PRODI**

- CPL1      Mahasiswa mampu mengidentifikasi dan menjelaskan dasar ilmu data yang meliputi sains dasar, matematika, statistika dan dasar pemrograman.
- CPL2      Mahasiswa mampu menerapkan teori, Teknik, dan alat bidang sains data untuk menyelesaikan permasalahan sederhana pada kehidupan sehari-hari.

## **CAPAIAN PEMBELAJARAN MATA KULIAH**

1. Mahasiswa mampu memahami operasi matriks dan vektor dengan numpy.
2. Mahasiswa mampu memahami proses pengolahan dataframe dengan pandas.
3. Mahasiswa mampu mengaplikasikan proses mengkoleksi data.
4. Mahasiswa mampu mengaplikasikan proses membersihkan noise data dan mencari data hilang.
5. Mahasiswa mampu menerapkan proses memperkaya data untuk mengatasi data hilang.
6. Mahasiswa mampu mengaplikasikan proses normalisasi dan standarisasi data.

# RENCANA PERKULIAHAN

Minggu Ke-	Pertemuan Ke-	Materi
I	1-2	Pendahuluan (KONTRAK KULIAH) Pengenal data wrangling dan fungsinya. The NumPy ndarray: array multidimensional. Fungsi universal pada numpy.
II	3-4	Numpy matriks. Numpy vector. Input dan output file array.
III	5-6	Pengenal struktur data dengan pandas. Indexing dataframe. Selection and filtering.
IV	7-8	Aritmatika dan data alignment. Pengurutan dan perangkatian. Summarizing dan komputasi statiska deskriptif

# RENCANA PERKULIAHAN

Minggu Ke-	Pertemuan Ke-	Materi
V	9-10	Membaca dan menulis data pada format teks. Data CSV dan XLS Data JSON Data XML Database
VI	11-12	Apa itu scraping dan bagaimana? Inspeksi elemen web Pengenalan dengan BeautifulSoup How to load page?
VII	13-14	Scraping tabel page tunggal. Scraping table multi page.
VIII	15-16	Ujian Tengah Semester

# RENCANA PERKULIAHAN

Minggu Ke-	Pertemuan Ke-	Materi
IX	17-18	Filtering data yang hilang. Memutuskan dan mengisi data yang hilang.
X	19-20	Proses-proses transformasi data Removing Duplicate Transformasi data menggunakan fungsi Identifikasi nilai data
XI	21-22	Diskritisasi and bucketing Deteksi data outlier Permutasi random sampling Manipulasi string
XII	23-24	Konsep data enriching Reordering dan pengurutan level Rangkuman statistic berdasarkan level Indexing dengan kolom dataframe Merger dua dataset Concatenating



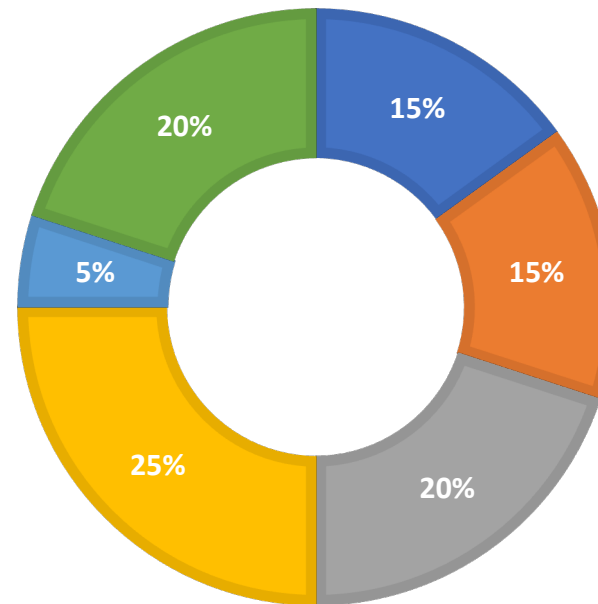
# RENCANA PERKULIAHAN

Minggu Ke-	Pertemuan Ke-	Materi
XII	25-26	Reshaping Pivoting “Long” to “Wide” format. Pivoting “Wide” to “Long” format.
XIV	27-28	Konsep normalisasi data Metode-metode normalisasi data Fungsi normalisasi data.
XV	29-30	Konsep standarisasi data Fungsi standarisasi data. Standarisari data pada python.
XVI	31-32	Ujian Akhir Semester

# PENILAIAN

## PENILAIAN MAHASISWA

■ Kuis 1 ■ Kuis 2 ■ UTS ■ UAS ■ Tugas ■ Praktikum



# REFERENSI

- Jacqueline Kazil, and Katharine Jarmul, 2016. Data Wrangling with Python: tips and tools to make your life easier. O'Reilly Media, Inc.
- Jaime Buelta, 2020. Python Automation Cookbook: 75 Python automation ideas for web scraping, data wrangling, and processing Excel, reports, emails, and more (2<sup>nd</sup> Edition). Packt Publishing Ltd. Birmingham.
- Wes McKinney, 2017. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython (2<sup>nd</sup> Edition). O'Reilly Media, Inc.

## ATURAN LAINNYA

- Tingkat kehadiran dosen didalam perkuliahan minimal 90%.
- Dosen mengajar sesuai dengan RPS.
- Mahasiswa yang dapat mengikuti UTS dan UAS apabila kehadiran tidak kurang dari 80%.
- Berpakaian rapi
- Bagi yang mengikuti secara daring Wajib On Cam.
- Wajib melaksanakan protocol Kesehatan
- Jaga Kesehatan ^\_^

# **RENCANA PEMBELAJARAN SEMESTER**



**IT Telkom  
Surabaya**

***Solution for The Nation***

**TERIMA  
KASIH**