

LAPORAN AKHIR
MAGANG & STUDI INDEPENDEN BERSERTIFIKAT

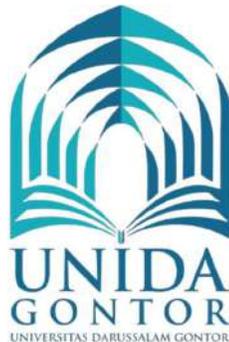
Data Science
Studi Independen
Di PT. Cendekia Data Andalan

Luqyana Nurul Husnina

432022618024

Nama Dosen Pendamping Program (DPP) :

Etistika Yuni Wijaya, S.Pd., M.Pd.



PROGRAM STUDI TEKNIK INFORMATIKA
UNIVERSITAS DARUSSALAM GONTOR

2024

LAPORAN PRAKTEK KERJA LAPANGAN

**Magang dan Studi Independen Bersertifikat (MSIB)
Di PT. Cendekia Data Andalan (Data Academy)**

Disusun oleh:

Luqyana Nurul Husnina 432022618024

Dipertanggungjawabkan di depan dosen penguji

Pada Selasa, 1 Juli 2025

**Menyetujui,
Dosen Penguji PKL**



Faisal Reza Pradhana, S.Kom., M.Kom

NIY. 160598

**Mengetahui,
Ketua Prodi Teknik Informatika UNIDA Gontor**



Dihin Muriyatmoko, S.ST., M.T.

NIY. 150489

Kata Pengantar

*Puji syukur ke hadirat Allah SWT atas limpahan rahmat dan karunia-Nya sehingga saya dapat menyelesaikan pelatihan ini dengan baik. Saya ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada pihak **Kampus Merdeka** dan **Data Academy** atas kesempatan yang diberikan untuk mengikuti kelas **Data Science**. Pelatihan ini telah memberikan pengalaman yang sangat berharga dalam memahami konsep, teknik, dan aplikasi data science yang relevan di era digital ini.*

Ucapan terima kasih juga saya sampaikan kepada seluruh instruktur dan tim pengajar yang telah membagikan ilmu dengan sabar dan profesional, serta memberikan bimbingan yang luar biasa sepanjang proses pembelajaran. Pengetahuan yang saya peroleh tidak hanya membantu saya dalam meningkatkan keterampilan teknis, tetapi juga membuka wawasan baru tentang pentingnya pengelolaan dan analisis data dalam berbagai bidang.

Tak lupa, saya haturkan terima kasih kepada rekan-rekan peserta yang telah menciptakan suasana belajar yang inspiratif dan kolaboratif. Diskusi dan kerja sama yang terjalin menjadi bagian penting dalam perjalanan pembelajaran ini.

Semoga ilmu yang didapatkan dapat bermanfaat dan menjadi langkah awal untuk terus berkembang di dunia data science.

Daftar Isi

Kata Pengantar	2
Daftar Isi	3
Daftar Gambar	4
Daftar Tabel	5
I. Gambaran Umum.....	6
A. Profil Perusahaan.....	6
B. Deskripsi Kegiatan.....	7
II. Aktivitas Bulanan.....	22
III. Penutup.....	26
A. Kesimpulan.....	26
B. Saran.....	26
Referensi.....	27
Lampiran.....	28

Daftar Gambar

<i>Gambar 1. 1 data sebelum di Cleaning</i>	19
Gambar 1. 2 Data Sesudah di Cleaning	19
Gambar 1. 3 Code Clusterring Methode	20
<i>Gambar 1. 4 Hasil Perbandingan Modelling dengan berbagai Methode</i>	20
Gambar 1. 5 Hasil Modelling menggunakan Elbow Method	20
<i>Gambar 1. 6 Dashboard Overview</i>	21
Gambar 1. 7 Dashboard Modelling	21
Gambar 1. 8 Dashboard Problem Solution	22

Daftar Tabel

Tabel 1. 1 Aktifitas dan Posisi Studi Independen	8
Tabel 1. 2 <i>Detail Program Studi Independen Data Academy</i>	8
Tabel 1. 3 Rincian Modul Pembelajaran Data Academy.....	9
Tabel 1. 4 Timeline Kegiatan di Data Academy	16
Tabel 1. 5 <i>Timeline Pengerjaan Tugas Akhir</i>	18

I. Gambaran Umum

A. Profil Perusahaan

PT. Cendekia Data Andalan (Data Academy) adalah *Learning Centre* untuk Professional Data. Data Academy menggabungkan antara Pendidikan, Industri, Pemerintahan dan Komunitas dalam mengembangkan kompetensi talenta Data Scientist & Artificial Intelligence (AI). Data Academy didirikan pada Januari 2020. Data Academy berfokus di area *Data Science* dan *Artificial Intelligence*. Dengan membangun ekosistem yang diberi nama Hexa Helix konsep, dimana Data Academy melakukan sinergi dengan beberapa pihak mulai dari Universitas, Institusi Pemerintahan, Korporasi, Principal, Profesional, dan Komunitas dalam menyiapkan talenta-talenta dibidang Data.

Perusahaan ini memiliki visi “*Nation Wide Leader as Learning Center for Digital Technology Specialty on Data Science & AI* “. Dengan beberapa misi yang diterapkan, yaitu :

1. Strengthening Synergy within Digital Ecosystem Stakeholder to reach mutual business benefits (Data Academy Hexa Helix).
2. Expanding Vertical and Horizontal across Industry, Education, Public segment Building Competency based Learning with
3. Building Competency based Learning with National or International standart.
4. Building Innovative Learning Platform to accelerate National wide market penetration.
5. Leading or Learning center for Digital Talent especially on Data Science for National Market.

DATA ACADEMY HEXA HELIX konsep menyediakan platform LINK and MATCH antara Dunia PENDIDIKAN dan Dunia INDUSTRI



B. Deskripsi Kegiatan

Data Science

Ilmu data (*data science*) adalah bidang interdisipliner yang menggunakan teknik dan metodologi untuk mengekstrak wawasan dari data terstruktur dan tak terstruktur. Ilmu data melibatkan pengumpulan, pembersihan, analisis, dan visualisasi data untuk membantu organisasi membuat keputusan yang tepat. Komponen utamanya meliputi pengumpulan data, pembersihan data, analisis, pembelajaran mesin, visualisasi data, interpretasi dan komunikasi, serta etika dan privasi. Ilmu data diterapkan di berbagai industri seperti perawatan kesehatan, keuangan, pemasaran, dan pendidikan untuk meningkatkan pengambilan keputusan, meningkatkan efisiensi operasional, dan mendorong inovasi. Kemajuan teknologi, seperti analisis big data dan komputasi awan, memungkinkan pemrosesan dan analisis kumpulan data yang lebih besar.

Kompetensi yang dikembangkan :

1. *Statistical Analysis*
2. *Programming Skills*
3. *Data Manipulation and Cleaning*
4. *Machine Learning*
5. *Data Visualization*
6. *Critical Thinking*
7. *Problem Solving*
8. *Communication Skills*
9. *Ethics and Data Privacy*

Kegiatan yang telah dilakukan selama program Data Science ialah menghadiri setiap pertemuan rutin dengan mentor, baik di sesi materi maupun mentoring. Dengan

tujuan membahas materi, memberikan umpan balik terhadap tugas serta mendiskusikan kemajuan proyek. Melakukan tugas kelompok untuk mereview dan meresume jurnal internasional dengan fokus pada problem klasifikasi, dan menyajikannya dalam bentuk presentasi. Mempelajari dan menerapkan teknik preprocessing data, implementasi model machine learning dan juga visualisasi data. Meningkatkan pemahaman tentang algoritma machine learning, pengelolaan proyek datascience dan keterampilan komunikasi serta kolaborasi dalam tim. Menghadapi tantangan dalam memahami konsep kompleks dan mencari solusi melalui referensi tambahan dan diskusi dengan teman.

Estimasi waktu penyampaian materi berlangsung selama 1 hingga 3 minggu tergantung dari bobot sks materi tersebut. Estimasi waktu tersebut merupakan gabungan dari pembelajaran dan pengaplikasian yang dipelajari dengan proyek yang dikerjakan. Setelah 1 modul selesai, maka peserta akan berpindah untuk mempelajari modul lainnya. Dan disetiap minggunya terdapat lecture dengan instruktur dan juga mentoring session.

Berikut gambaran singkat dari desain Studi Independen Bersertifikat Kampus Merdeka Data Academy yang mencakup durasi dan tipe aktivitas :

Tabel 1. 1 Aktifitas dan Posisi Studi Independen

Durasi Aktifitas	900 jam
Jumlah SKS	20 SKS
Tipe Aktifitas	Pembelajaran Mandiri (Self Study), Diskusi Grup (Peer-to Peer Discussion), Pendampingan (Mentoring), Kuliah Umum (Lecture), Tugas Individu (Individual Assignments), Tugas Kelompok (Capstone Project)
Lokasi Aktifitas	Belajar dari Rumah (<i>Learning from Home</i>)

Berikut detail program beserta target setiap modul dari program Data Academy yang telah diajarkan.

Tabel 1. 2 Detail Program Studi Independen Data Academy

Detail Program Studi Independen	
Nama Aktifitas	Data Academy – Data Science Academy

Durasi Program	6 September – 20 Desember 2024
SKS	20
Tipe	Daring
Lokasi	Daring
Deskripsi	Program <i>Data Academy</i> untuk kelas <i>Data Science</i> dirancang untuk membekali peserta dengan keterampilan analisis data, pemrograman, dan <i>machine learning</i> melalui kurikulum komprehensif yang mencakup statistik, Python, serta visualisasi data menggunakan alat seperti Tableau dan Google Colab. Dengan bimbingan instruktur berpengalaman dan proyek nyata, peserta akan belajar mengolah, menganalisis, dan memvisualisasikan data untuk menghasilkan wawasan yang relevan. Program ini juga menekankan kolaborasi dan pembelajaran berbasis kasus, sehingga setelah selesai, peserta mampu membuat model prediktif dan mempresentasikan hasil analisis secara profesional, membuka peluang karier di bidang <i>Data Science</i> .

Tabel 1. 3 Rincian Modul Pembelajaran Data Academy

Modul 1. Pengantar Data Science	
Detail Pembelajaran	<ul style="list-style-type: none"> • Pengantar Data Science • Data Mining dan Data Modelling • Pengantar Big Data Science • Perkembangan dan Peran Big Data
Target Hasil Pembelajaran	<ul style="list-style-type: none"> • Mampu menjelaskan konsep dasar dan konteks dalam bidang Data Science, Data Modelling, Data Mining dan pengaplikasiannya di industri. • Mampu menjelaskan konsep dasar big data science. • Mampu menjelaskan kebutuhan dan prospek big data science.

Durasi Pembelajaran	100 jam
SKS	2
Bentuk Pembelajaran	<ul style="list-style-type: none"> • Kelas Webinar Online (Ceramah, Tutorial, Diskusi) • Tugas melalui Learning Management System (LMS) • Mentoring dan Konsultasi Online • Penggunaan Perangkat Lunak Microsoft Excel, Google Spreadsheet.
Modul 2. Bussiness dan Data Understanding	
Detail Pembelajaran	<ul style="list-style-type: none"> • Analytics Design Thinking. • Proses Analisa untuk Pemecahan Masalah • Tipe Data Analitik • Perbedaan Analytik vs Analisis
Target Hasil Pembelajaran	<ul style="list-style-type: none"> • Mampu menjelaskan pentingnya berpikir analitik. • Mampu menjelaskan proses analytical problem-solving. • Mampu menjelaskan tipe Data Analytics. • Mampu menjelaskan perbedaan Analysis vs Analytics.
Durasi Pembelajaran	<ul style="list-style-type: none"> • 100 jam
SKS	<ul style="list-style-type: none"> • 2
Bentuk Pembelajaran	<ul style="list-style-type: none"> • Kelas Webinar Online (Ceramah, Tutorial, Diskusi) • Tugas melalui Learning Management System (LMS) • Mentoring dan Konsultasi Online • Penggunaan Perangkat Lunak Data Visualisasi Tableau
Modul 3. Pengumpulan dan Validasi Data	
Detail Pembelajaran	<ul style="list-style-type: none"> • Pengumpulan Data (Identifikasi Sumber Data, Pengumpulan Data Primer dan Sekunder), Evaluasi Kualitas dan Relevansi Data.

	<ul style="list-style-type: none"> • Menelaah Data (Eksplorasi Data Awal, Identifikasi Pola dan Tren Awal), Analisis Deskriptif Data) • Memvalidasi Data (Pemeriksaan Kesalahan dan Anomali, Penanganan Missing Values, Verifikasi Konsistensi dan Kualitas Data)
Target Hasil Pembelajaran	<ul style="list-style-type: none"> • Mengerti dan memahami konteks mengumpulkan data dan validasi data. • Mampu mengidentifikasi kebutuhan data berdasarkan objective. • Mampu mengumpulkan dan validasi data.
Durasi Pembelajaran	<ul style="list-style-type: none"> • 120 jam
SKS	<ul style="list-style-type: none"> • 2
Bentuk Pembelajaran	<ul style="list-style-type: none"> • Kelas Webinar Online (Ceramah, Tutorial, Diskusi) • Tugas melalui Learning Management System (LMS) • Mentoring dan Konsultasi Online • Penggunaan Perangkat Lunak Microsoft Excel, Talend Studio (versi terbaru).
Modul 4. Dasar Statistika untuk Data Science	
Detail Pembelajaran	<ul style="list-style-type: none"> • Dasar-dasar Statistika untuk Data Science • Pengenalan MYSQL Database dengan Implementasi Statistik. • Pengantar, Perancangan, dan Implementasi Data Modelling
Target Hasil Pembelajaran	<ul style="list-style-type: none"> • Mampu mengaplikasikan konsep statistika dalam analisis data. • Mampu merancang dan mengembangkan model data yang efisien dan efektif.
Durasi Pembelajaran	<ul style="list-style-type: none"> • 60 jam
SKS	<ul style="list-style-type: none"> • 2
Bentuk Pembelajaran	<ul style="list-style-type: none"> • Kelas Webinar Online (Ceramah, Tutorial, Diskusi) • Tugas melalui Learning Management

	<p>System (LMS)</p> <ul style="list-style-type: none"> • Mentoring dan Konsultasi Online • Penggunaan Perangkat Lunak Microsoft Excel, Google Spreadsheet.
Modul 5. Dasar Python Untuk Data Science	
Detail Pembelajaran	<ul style="list-style-type: none"> • Dasar Python untuk Data Science. • Pengolahan Data dan Eksploratory Data Analytic dengan Python
Target Hasil Pembelajaran	<ul style="list-style-type: none"> • Mampu menjelaskan dasar dasar dalam menggunakan bahasa pemrograman Python untuk Data Science. • Mampu mengolah dan mengeksplorasi data menggunakan teknik dan perangkat lunak yang relevan.
Durasi Pembelajaran	<ul style="list-style-type: none"> • 120 jam
SKS	<ul style="list-style-type: none"> • 2
Bentuk Pembelajaran	<ul style="list-style-type: none"> • Kelas Webinar Online (Ceramah, Tutorial, Diskusi) • Tugas melalui Learning Management System (LMS) • Mentoring dan Konsultasi Online • Penggunaan Perangkat Lunak Python (versi terkini); Integrated Development Environment (IDE) seperti PyCharm, Jupyter Notebook, atau Visual Studio Code, Python (versi terkini); IDE seperti PyCharm, Jupyter Notebook, atau Visual Studio Code; Library Python (pandas, numpy, matplotlib, seaborn)
Modul 6. Pemodelan Data dan Machine Learning	
Detail Pembelajaran	<ul style="list-style-type: none"> • Pengantar Machine Learning • Proses dan Teknik Pemodelan Machine Learning • Supervised Learning (regresi, klasifikasi) dan unsupervised Learning (clustering, reduksi dimensi) • Evaluasi model dan pemilihan fitur

	<ul style="list-style-type: none"> • Algoritma Machine Learning (regresi linier, regresi logistik, k-NN, SVM, Naive Bayes, Decision Trees, Random Forest, K Means, PCA) • Pengantar Deep Learning dan Neural Networks, - Text Mining dan Natural Language Processing (NLP) • Feature extraction (Bag of Words, TF IDF), Algoritma NLP (Naive Bayes, k-NN, SVM, LSTM)
Target Hasil Pembelajaran	<ul style="list-style-type: none"> • Mampu mengaplikasikan teknik Machine Learning dalam menyelesaikan masalah Data Science. • Mampu mengaplikasikan teknik Advanced Data Science dalam berbagai kasus di dunia nyata.
Durasi Pembelajaran	<ul style="list-style-type: none"> • 120 jam
SKS	<ul style="list-style-type: none"> • 3
Bentuk Pembelajaran	<ul style="list-style-type: none"> • Kelas Webinar Online (Ceramah, Tutorial, Diskusi) • Tugas melalui Learning Management System (LMS) • Mentoring dan Konsultasi Online • Penggunaan Perangkat Lunak Microsoft Excel, Google Spreadsheet, Google Colab, Library Python (pandas, numpy, matplotlib, seaborn)
Modul 7. Data Interpretation	
Detail Pembelajaran	<ul style="list-style-type: none"> • Pengenalan Aplikasi Tableau Desktop • Terminologi Data Interpretasi dan Visualisasi • Kode-kode Visual dalam Aplikasi Data Visualisasi • Visualisasi Data Geografis • Jenis-jenis Data • Membuat Grafik • Membuat Dashboard • Tahapan Membuat Storytelling

Target Hasil Pembelajaran	<ul style="list-style-type: none"> • Mampu mengetahui lingkungan aplikasi perangkat lunak data visualisasi. • Mampu mengenal jenis data dan menghubungkannya ke aplikasi data visualisasi. • Mampu mengeksplorasi dan menganalisa data menggunakan aplikasi data visualisasi. • Mampu memvisualisasikan data geografis. • Mampu melakukan kalkulasi dalam aplikasi data visualisasi. • Mampu membuat grafik dengan efektif dan efisien. • Kemampuan dalam membuat dasbor. • Mampu membuat story telling.
Durasi Pembelajaran	<ul style="list-style-type: none"> • 100 jam
SKS	<ul style="list-style-type: none"> • 2
Bentuk Pembelajaran	<ul style="list-style-type: none"> • Kelas Webinar Online (Ceramah, Tutorial, Diskusi) • Tugas melalui Learning Management System (LMS) • Mentoring dan Konsultasi Online • Penggunaan Perangkat Lunak Tableau
Modul 8. Soft Skill	
Detail Pembelajaran	<ul style="list-style-type: none"> • Communication Skill • Public Speaking • Personal Branding • Presentation Skill • Resume and Interview Preparation
Target Hasil Pembelajaran	<ul style="list-style-type: none"> • Mampu menerapkan sikap teliti dan kritis dalam mempersiapkan diri sebagai Data Scientist. • Mampu menerapkan kemampuan komunikasi dan presentasi yang baik. • Mampu mengimplementasikan manajemen waktu dan kerjasama tim. • Mampu mempersiapkan resume dan wawancara memasuki dunia kerja.

Durasi Pembelajaran	<ul style="list-style-type: none"> • 60 jam
SKS	<ul style="list-style-type: none"> • 2
Bentuk Pembelajaran	<ul style="list-style-type: none"> • Kelas Webinar Online (Ceramah, Tutorial, Diskusi) • Tugas melalui Learning Management System (LMS) • Mentoring dan Konsultasi Online • Penggunaan Perangkat Lunak Microsoft Word, Microsoft Presentation, atau Aplikasi Online Presentation.
Modul 9. Final Project	
Detail Pembelajaran	<ul style="list-style-type: none"> • Merumuskan Permasalahan Studi Kasus • Pembuatan Project untuk Mengolah dan Mentransformasi Data Skala Besar menggunakan Tools Data Visualisasi • Pembuatan Interactive Presentation • Mempresentasikan hasil analisis
Target Hasil Pembelajaran	<ul style="list-style-type: none"> • Mampu untuk mengimplementasikan ilmu yang sudah dipelajari sebelumnya dalam karya berupa aplikasi. • Mampu berkreatifitas dan mencetuskan ide untuk membuat karya yang bernilai. • Mampu mengetahui dasar-dasar bagaimana memvalidasi sebuah ide sebelum diterapkan. • Mampu bekerja sama dan berkolaborasi dengan teman satu tim (kerjasama tim). • Menganalisis dan identifikasi masalah. • Mencari solusi dari permasalahan. • Mencari ide dan berpikir kreatif. • Mengeksekusi solusi dengan efektif dan efisien. • Mengidentifikasi dan mengelola proyek. • Menyelesaikan proyek tepat waktu. • Presentasi proyek.
Durasi Pembelajaran	<ul style="list-style-type: none"> • 120 jam
SKS	<ul style="list-style-type: none"> • 2

Bentuk Pembelajaran	<ul style="list-style-type: none"> • Kerja Kelompok • Mentoring dan Konsultasi Online • Penggunaan Perangkat Lunak Data Visualisasi, Microsoft Presentation, atau Aplikasi Online Presentation Total Pembelajaran
---------------------	--

Pada jadwal kerja, berikut merupakan timeline dari kegiatan pada studi independen di PT. Cendekia Data Andalan :

Tabel 1. 4 Timeline Kegiatan di Data Academy

No	Kegiatan	Minggu ke-														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	Pengantar Data Science															
2	Bussiness dan Data Understanding															
3	Pengumpulan dan Validasi Data															
4	Dasar Statistika untuk Data Science															
5	Dasar Python untuk Data Science															
6	Pemodelan Data dan Machine Learning															
7	Data Interpretation															

8	Soft Skill														
9	Final Project														

Lingkup Pembelajaran dan Proyek

Selama berlangsungnya program, peserta diberikan beberapa tugas yang sesuai dengan materi yang diberikan. Tugas yang diberikan dapat berbentuk individu dan kelompok. Lingkup tersebut meliputi :

A. Tugas Berkala

Tugas yang diberikan dalam program pembelajaran Data Science ini melibatkan berbagai aspek pembelajaran yang terintegrasi dan berfokus pada pemahaman teori, implementasi, serta kolaborasi praktis. Pertama, peserta diminta melakukan tugas *business understanding* berupa mereview jurnal internasional untuk memahami konsep dasar, metodologi, dan aplikasi dari data science dalam berbagai konteks bisnis atau penelitian.

Selanjutnya, tugas pengumpulan dan visualisasi data melibatkan pencarian jurnal internasional yang menerapkan materi yang telah diajarkan, guna mengeksplorasi bagaimana teknik pengolahan dan visualisasi data digunakan dalam penelitian. Pada tahap berikutnya, peserta mengerjakan tugas individual berupa praktikal penggunaan Python, khususnya pada materi dasar Python untuk *data science*, seperti manipulasi data, analisis statistik, dan pembuatan visualisasi sederhana.

Kemudian, dalam kerja kelompok, peserta mengembangkan pemodelan algoritma *machine learning* untuk kasus deteksi penyakit hati, mencakup proses mulai dari eksplorasi data, *feature engineering*, hingga evaluasi model. Terakhir, peserta juga bekerja secara berkelompok untuk memvisualisasikan data kemiskinan, menyajikan wawasan berbasis data secara kreatif dan informatif menggunakan alat visualisasi seperti Tableau atau Matplotlib, untuk menggambarkan dinamika sosial-ekonomi.

Setiap tahapan ini dirancang untuk memperkuat pemahaman teori dan keterampilan praktis, sekaligus membangun kemampuan kolaborasi dalam tim.

B. Pelaksanaan tugas akhir Data Science

Pada pengerjaan tugas akhir ini, dibuatlah timeline agar memudahkan mahasiswa dalam menyelesaikan project, yaitu sebagai berikut :

Tabel Timeline Pengerjaan Tugas Akhir :

Tabel 1. 5 Timeline Pengerjaan Tugas Akhir

No	Deadline	To Do List
1	28 November 2024	<ul style="list-style-type: none">• Bussiness Understanding• Pengumpulan Data• Menentukan Algoritma Pemodelan Data
2	5 Desember 2024	<ul style="list-style-type: none">• Menerapkan algoritma pada data• Membuat story board
3	12 Desember 2024	<ul style="list-style-type: none">• Memvisualisasikan data yang sudah di clusterring• Pembuatan dashboard visualisasi
4	17 Desember 2024	<ul style="list-style-type: none">• Presentasi project akhir

C. Pencapaian Tugas Akhir

a. Understading Bussiness Problem

Dalam project ini kami mengambil kasus mengenai pengelolaan sampah di berbagai provinsi di indonesia. Kami menganalisis bagaimana di setiap provinsi indonesia memastikan dalam pengelolaan sampah yang ada. Dengan mempertimbangkan beberapa faktor pemicu. Seperti, distribusi jumlah tempat pembuangan sampah (TPS), kapasitas TPS, jumlah sampah berbahaya di setiap provinsi, serta tingkat pengelolaan sampah di masing-masing wilayah. Sehingga kami merangkumnya dalam bentuk 5W1H.

b. Data Cleaning

Setelah terkumpulnya data yang dibutuhkan, kami melakukan cleansing pada data. Yang dimana mebersihkan data dari nilai-nilai yang tidak lengkap. Dan menghilangkan fitur-fitur yang tidak diperlukan serta menggabungkan data-data tersebut menjadi satu. Untuk referensi data yang kami ambil bersumber dari website satu data ditjen PSLB3.

Provinsi	Kategori Sampah	Subkategori Sampah	Sampah Dikumpulkan (Kg)	Sampah Dikelola (Kg)	Sampah Sisa (Kg)
1 Aceh	Kertas	Kardus	21784.48	18555.25	3129.23
2 Aceh	Plastik	1-PEt	15515.55	12832.55	2683
3 Aceh	Organik	Lainnya	1195	0	1195
4 Aceh	Logam	Besi	80027	0	80027
5 Aceh	Sampah Lainnya	Kaca dan Beling	59003	0	59003
6 Aceh	Organik	Sisa Makanan	44000	0	44000
7 Aceh	Sampah Lainnya	Tekstil dan Produk Tekstil	300	0	300
8 Aceh	Sampah Lainnya	Karet dan kulit	300	0	300
9 Bali	Plastik	1-PEt	1120541.36	679398.3	450543.06
10 Bali	Kertas	Kardus	327777.86	208228.29	129549.57
11 Bali	Logam	Besi	242726.28	25650.49	217075.79
12 Bali	Sampah Lainnya	Kaca dan Beling	41122.25	52217.42	8193.83
13 Bali	Sampah Lainnya	Karet dan kulit	19082.37	18316.6	765.77
14 Bali	Organik	Sampah Mudah Terurai	15169.7	15162	7.7
15 Bali	Organik	Sisa Makanan	24024	13310	10714
16 Bali	Organik	Lainnya	5746	3263	2483
17 Bali	Sampah Lainnya	Limbah B3	3603.3	3393.5	209.8
18 Bali	Sampah Lainnya	Tekstil dan Produk Tekstil	1947.5	1930.8	166.7
19 Bali	Sampah Elektronik	Televisi (TV)	1214.9	1181.5	33.4
20 Bali	Sampah Lainnya	Limbah B3	1665.2	1059.3	605.9
21 Bali	Kertas	Duploak	48500	0	48500
22 Banten	Plastik	1-PEt	23628.28	10482	13146.28
23 Banten	Kertas	Kardus	18465.45	10383.5	8081.95
24 Banten	Organik	Sisa Makanan	4812	0	4812
25 Banten	Logam	Besi	4007.35	3175.45	831.9
26 Banten	Sampah Lainnya	Kaca dan Beling	1731	1320	411
27 Banten	Organik	Lainnya	88000	55200	32800
28 Banten	Organik	Sampah Mudah Terurai	17000	24000	39000
29 Banten	Sampah Elektronik	Televisi (TV)	13305	13145	160
30 Banten	Sampah Lainnya	Karet dan kulit	5480	5480	0
31 Bengkulu	Kertas	Kardus	20052.4	10018.6	10033.8
32 Bengkulu	Sampah Lainnya	Limbah B3	3562.6	3113.1	449.5
33 Bengkulu	Sampah Lainnya	Limbah B3	3213.4	0	3213.4
34 Bengkulu	Plastik	1-PEt	1156.2	694.4	461.8
35 Bengkulu	Kertas	Duploak	33020	0	33020
36 Bengkulu	Logam	Besi	25740	20700	5040
37 Bengkulu	Plastik	7-Plastik Lainnya	13980	0	13980

Gambar 1. 1 data sebelum di Cleaning

Gambar 1. 2 Data Sesudah di Cleaning

c. Data Modelling

Dalam upaya memahami efisiensi pengelolaan sampah di setiap provinsi, dilakukan *clustering* berdasarkan rasio sampah yang tidak terolah. Pemodelan ini bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan tingkat efisiensi pengelolaan sampahnya, sehingga dapat diidentifikasi area yang memerlukan perhatian khusus.

Proses modeling dimulai dengan metode *Elbow Method*, yang digunakan untuk menentukan jumlah cluster optimal. Grafik di bawah ini menunjukkan bahwa jumlah cluster optimal adalah 3,

karena pada titik ini terjadi penurunan signifikan dalam nilai WSS (*Within-Cluster Sum of Squares*).

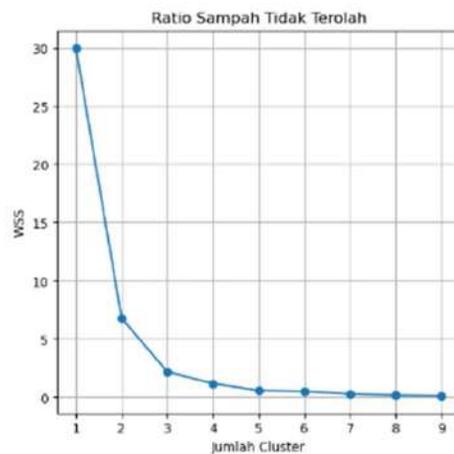
Selanjutnya, dilakukan evaluasi dengan membandingkan beberapa algoritma *clustering*, yaitu Agglomerative Clustering, DBSCAN, Gaussian Mixture Model (GMM), dan K-Means. Berdasarkan hasil evaluasi, algoritma Agglomerative Clustering dan K-Means memberikan nilai *Silhouette Score* tertinggi sebesar 0.70, menunjukkan bahwa pembagian cluster memiliki kualitas yang baik dalam memisahkan data antar cluster. Dalam kasus ini, algoritma yang digunakan adalah K-Means karena efisiensinya dalam hal komputasi. Hasil *clustering* kemudian diaplikasikan pada setiap provinsi dengan label yang sesuai dengan karakteristik rasio sampah tidak terolahnya.

```
# K-Means Clustering
kmeans = KMeans(n_clusters=optimal_clusters_dict[col], random_state=42)
kmeans_labels = kmeans.fit_predict(X)
kmeans_score = evaluate(X, kmeans_labels)
```

Gambar 1. 3 Code Clustering Methode

Feature Name	Model	Silhouette Score	Labels
Ratio Sampah Tidak Terolah	Agglomerative	0.70	2 0 0 1 2 2 0 2 1 1 1 2 1 1 1 0 0 1 2 0 0 2 1 2 1 1 2 0 2 0
	DBSCAN	-1.00	0 0
	GMM	0.67	2 1 1 2 0 0 1 0 2 2 2 0 2 2 2 1 1 2 0 1 1 0 2 0 2 2 0 1 0 1
	K-Means	0.70	0 1 1 2 0 0 1 0 2 2 2 0 2 2 2 1 1 2 0 1 1 0 2 0 2 2 0 1 0 1

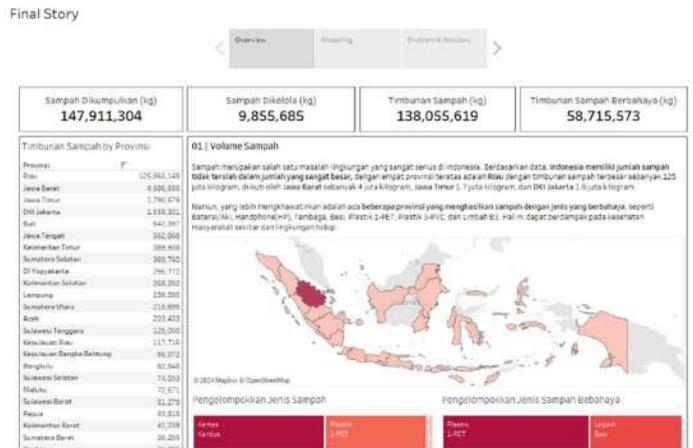
Gambar 1. 4 Hasil Perbandingan Modelling dengan berbagai Methode



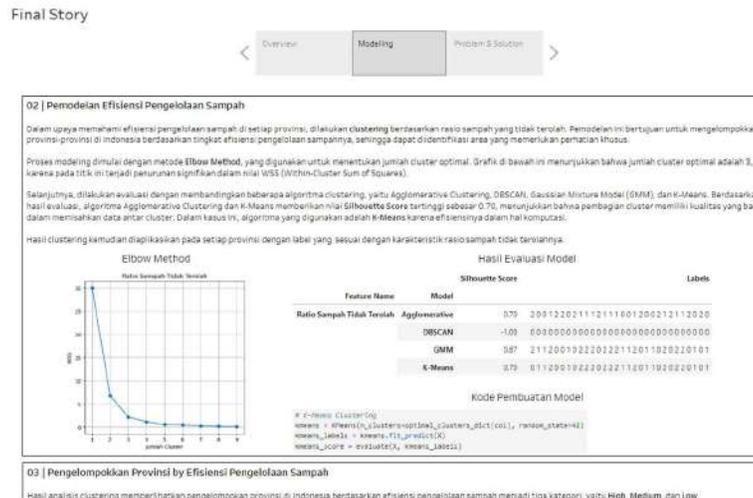
Gambar 1. 5 Hasil Modelling menggunakan Elbow Method

d. Data Visualization

Untuk memudahkan penyampaian informasi kepada audiens. Kami membuat visualisasi informasi tersebut dengan menarik dengan menarik agar mudah dipahami. Berikut hasil visualisasi kami :



Gambar 1. 6 Dashboard Overview



Gambar 1. 7 Dashboard Modelling



Gambar 1. 8 Dashboard Problem Solution

II. Aktivitas Bulanan

Berikut pemaparan kegiatan pembelajaran yang dilaksanakan dalam instansi per bulannya :

Bulan	Kegiatan
1	Pada bulan pertama di kelas <i>Data Science</i> , beberapa aktivitas dilakukan, termasuk pendampingan, pemberian tugas, pemecahan masalah, dan pengembangan keterampilan. Mentor berperan aktif dalam memberikan umpan balik, menjawab pertanyaan, dan berkolaborasi dengan Program Manajemen Data (DPP) untuk mendukung pembelajaran dan kemajuan tugas. Tugas-tugas tersebut meliputi peninjauan dan penyajian jurnal dengan kasus masalah yang dapat diklasifikasikan, pengumpulan data, pra-pemrosesan, identifikasi pelanggan, perkiraan harga, dan prediksi. Tantangannya adalah memahami konsep-konsep yang disajikan dalam makalah, yang diajarkan melalui berbagai materi seperti dokumen, buku, atau tutorial video. Pengembangan kompetensi sangat penting untuk mengidentifikasi masalah data dan menemukan solusi yang efektif.

2	<p>Aktivitas pendampingan pada bulan kedua dalam keilmuan data melibatkan pengembangan pemahaman konsep seperti pengumpulan dan validasi data, pemrograman Python untuk ilmu data, serta pemodelan data dan pembelajaran mesin. Sesi-sesi ini melibatkan peninjauan tugas-tugas sebelumnya, evaluasi hasil analisis data, dan diskusi solusi untuk masalah. Fokus utamanya adalah pada praproses data, penerapan model pembelajaran mesin, pemahaman proses penambangan data, dan teknik transformasi data. Beberapa keterampilan utama meliputi pemahaman algoritma pembelajaran mesin, pembelajaran strategi pemblokiran waktu, dan pemahaman bisnis.</p> <p>Kompetensi yang dikembangkan dalam fase-fase ini meliputi praproses data, penerapan pembelajaran mesin, visualisasi data, dan pengelolaan proyek ilmu data. DPP (Departemen Pengembangan Proyek) bertanggung jawab untuk mengoordinasikan pengembangan pengetahuan ilmu data untuk proyek-proyek tertentu, memastikan bahwa kumpulan data dan hasil dari area-area ini dapat digunakan untuk pengembangan proyek. Tujuannya adalah untuk menyediakan data alur kerja yang efektif dan panduan yang tepat untuk penerapan algoritma pembelajaran mesin.</p>
3	<p>pada bulan ketiga Kegiatan pendampingan sudah terstruktur dan difokuskan pada pengenalan materi baru dan peningkatan proyek mahasiswa. Kegiatan ini meliputi interpretasi data, pembelajaran mesin, dan pemodelan data yang efektif. Mentor memberikan saran terperinci untuk mengatasi tantangan teknis dan meningkatkan kualitas proyek. Fokusnya adalah pada pemahaman model data yang kompleks, referensi berbagai sumber, penerapan algoritma pembelajaran mesin, pengembangan kreativitas dalam menciptakan grafik yang menarik secara visual, dan pengembangan kompetensi dalam interpretasi data, pembelajaran mesin, pemodelan data, dan visualisasi. Proyek akhir akan diselesaikan secara kolaboratif, yang melibatkan pengumpulan data, visualisasi data, dan pengambilan keputusan. Tujuannya adalah untuk meningkatkan kualitas proyek melalui komunikasi dan visualisasi yang efektif.</p>

4	<p>Pada bulan ini, fokus utama adalah penyelesaian final project. Aktivitas mentoring intensif dilakukan melalui pertemuan yang diadakan secara berkala dengan mempresentasikan hasil kerja kelompok yaitu dari presentasi pertama di tanggal 26 november 2024 hingga presentasi terakhir pada tanggal 17 desember 2024 untuk memastikan setiap peserta berada di jalur yang tepat. Mentor memberikan bimbingan khusus terkait desain proyek, analisis data, dan visualisasi hasil dalam bentuk grafik dan story dashboard. Selain itu, bagi peserta yang membutuhkan bantuan teknis atau masukan strategis dalam menyelesaikan proyek dapat ditanyakan melalui grup whatsapp kelas yang telah disediakan.</p> <p>Pengerjaan final project ini mencakup menentukan studi kasus sesuai minat dan kebutuhan. Lalu mengintegrasikannya dengan dataset dari beberapa sumber terpercaya. Untuk menyelesaikan permasalahannya, digunakanlah algoritma machine learning yang sesuai. Setelah itu dilaporkan hasilnya dengan cara presentasi kelompok secara visual dalam bentuk tableau dan python.</p> <p>Tantangan yang saya hadapi dalam menyelesaikan tugas ialah</p> <ul style="list-style-type: none"> • sulitnya mencari dataset dengan fitur yang dibutuhkan dan preprocessing pada data yang tidak dibutuhkan. Untuk menyelesaikannya saya mencari beberapa panduan dari beberapa studi kasus mengenai Langkah-langkah yang dilakukan untuk membersihkan data. • Saya mengalami sedikit keesulitan dalam menentukan model machine learning yang cocok untuk diaplikasikan dengan data. Untuk menyelesaikannya, saya mencari beberapa panduan mengenai pemilihan algoritma berdasarkan jenis data dan tujuan proyeknya. • Saya mengalami kesulitan dalam manajemen waktu untuk menyelesaikan tugas yang diberikan. Dikarenakan kurangnya pemahaman saya, sehingga saya membutuhkan waktu yang lebih lama dalam penyelesaiannya. Untuk mengatasinya saya membuat timeline untuk setiap tahapan proyek. <p>Pengembangan potensi yang saya rasakan selama masa penyelesaian projek akhir ini adalah saya merasa adanya peningkatan dalam mengidentifikasi masalah sehingga dapat</p>
---	---

	<p>menentukan data apa saja yang dibutuhkan untuk menyelesaikan permasalahan tersebut. saya juga merasa adanya peningkatan dalam pemilihan model algoritma machine learning sesuai dengan data yang ada. Saya juga merasa adanya peningkatan dalam pemahaman untuk memvisualisasikan data dengan grafik yang cocok dengan datanya.</p>
--	--

III. Penutup

A. Kesimpulan

Program ini telah membantu saya dalam :

- **Peningkatan Keterampilan**
Program ini telah membantus saya dalam meningkatkan kemampuan saya mengenai identifikasi masalah, pemilihan model algoritma machine learning, dan memvisualisasikannya dalam bentuk grafik.
- **Pengalaman Berharga**
Program ini merupakan pengalaman berharga bagi saya, dengan mengikuti program ini saya tidak hanya merasa pada peningkatan keterampilan saja, melainkan membuka wawasan saya akan pentingnya pengelolaan dan analisis data.
- **Kolaborasi dan Diskusi**
Program ini mengajarkan kepada saya bahwasannya suasana belajar yang inspiratif dan kolaboratif merupakan faktor penting dalam pembelajaran sehingga dapat membantu saya dalam menyelesaikan tugas akhir saya.

B. Saran

Saran yang saya dapat sampaikan ialah diadakannya event seperti halnya kegiatan networking dan job fair dengan perusahaan-perusahaan yang mencari telenta di bidang data science, sehingga para peserta dapat menjajaki peluang karir.

Referensi

<https://kampusmerdeka.kemdikbud.go.id/>

<https://dataacademy.co.id/about/>

*Safitra, Muhammad Fakhrol, Muharman Lubis, Tien Fabrianti Kusumasari, and Deyana Prastika Putri. 2024. "Advancements in Artificial Intelligence and Data Science: Models, Applications, and Challenges." *Procedia Computer Science* 234: 381–88. doi:10.1016/j.procs.2024.03.018.*

Lampiran

A. Letter of Acceptance



Nomor : 556/CDA/LOA-MSIB7/VIII-2024
 Lampiran : Satu berkas
 Hal : Penerimaan Mahasiswa Peserta (Studi Independen Bersertifikat) Program MSIB
 Kampus Merdeka

LETTER OF ACCEPTANCE

Saya yang bertanda tangan di bawah ini :

Nama Lengkap : Rahmah, S.Sos.
 Jabatan : Digital Account Manager
 Nama Perusahaan/ Organisasi : PT Cendekia Data Andalan

Selaku penanggungjawab Program Magang dan Studi Independen Bersertifikat (MSIB) Kampus Merdeka Angkatan 7 periode tahun 2024, dengan ini menyatakan bahwa nama-nama terlampir merupakan peserta program **Studi Independen** di PT Cendekia Data Andalan dengan pelaksanaan pada **6 September – 31 Desember 2024**.

Demikian surat ini kami sampaikan sebagai kelengkapan syarat administrasi program MSIB Angkatan 7 periode tahun 2024 dan dapat dipergunakan sebagaimana mestinya.

Jakarta, 6 September 2024

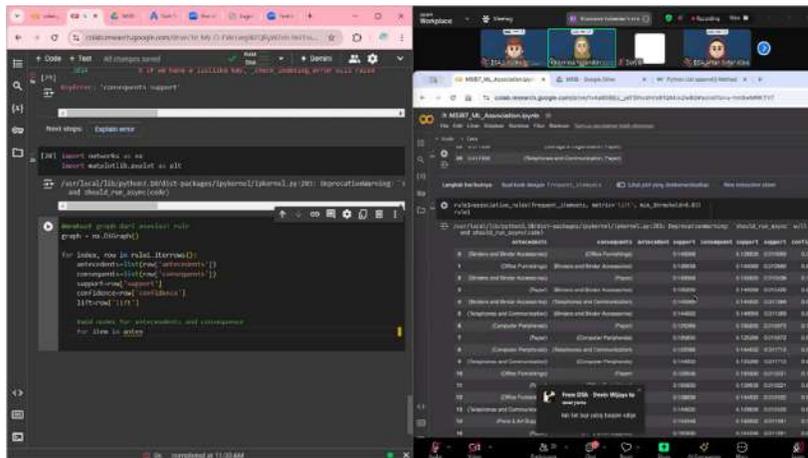
Rahmah, S.Sos
 Digital Account Manager

8	9542809	Eusggleton Nathan	S1	Universitas Esa Unggul	Teknik Informatika	Data Analyst Academy	Studi Independen	6 September 2024 - 31 Desember 2024	Prov. D.K.I. Jakarta
9	9646140	DERLYA MARDINAH	S1	Universitas Adhika	Informatika	Data Science Academy	Studi Independen	6 September 2024 - 31 Desember 2024	Prov. Sumatera Barat
10	10679603	GUSTI RAMADHANI	D4	Politeknik Negeri Stajawa	Manajemen Informatika	Data Analyst Academy	Studi Independen	6 September 2024 - 31 Desember 2024	Prov. Sumatera Selatan
11	9500808	LUQYANA NURUL HUSNINA	S1	Universitas Darussalam Gontor	Teknik Informatika	Data Science Academy	Studi Independen	6 September 2024 - 31 Desember 2024	Prov. Jawa Barat
12	11704571	TIKTA SAMUDERA RAMADHANI	S1	Universitas Pendidikan Indonesia	Sistem Informasi Kelautan	Data Science Academy	Studi Independen	6 September 2024 - 31 Desember 2024	Prov. Jawa Barat
13	9696880	DAVINA DWI ANGELICA	S1	Universitas Pancasila	Teknik Informatika	Data Science Academy	Studi Independen	6 September 2024 - 31 Desember 2024	Prov. D.K.I. Jakarta
14	8724351	Mohammad Faqih	S1	Universitas Jember	Sistem Informasi	Data Science Academy	Studi Independen	6 September 2024 - ...	Prov. Jawa Timur

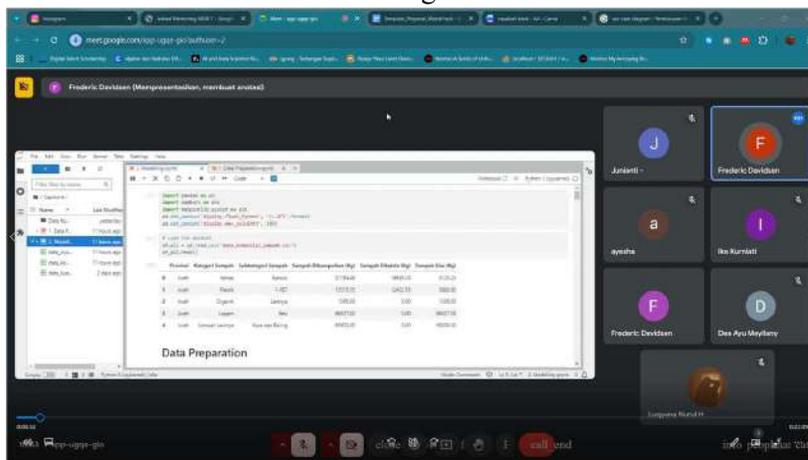
B. Dokumentasi Live class Dasar Python untuk Data Science



C. Dokumentasi Live Class Pemodelan Data dan Machine Learning



D. Dokumentasi Live Internal Meeting/ Discussion Team



I. Sertifikat Kompetensi



J. Transkrip Nilai Magang

