

**LAPORAN AKHIR**  
**MAGANG & STUDI INDEPENDEN BERSERTIFIKAT**  
**DATA SCIENTIST**  
**DI MYEDUSOLVE - PT DWI INTI PUTRA**

Diajukan untuk memenuhi persyaratan kelulusan  
Program MSIB MBKM

oleh:

Najwa Amelia Qorry 'Aina / 5027201001



**DEPARTEMEN TEKNOLOGI INFORMASI**  
**INSTITUT TEKNOLOGI SEPULUH NOPEMBER**  
**2022**

**Lembar Pengesahan Departemen Teknologi Informasi**  
**Institut Teknologi Sepuluh Nopember**  
**DATA SCIENTIST**  
**Di MyEduSolve - PT Dwi Inti Putra**

oleh:

Najwa Amelia Qorry 'Aina / 5027201001

disetujui dan disahkan sebagai  
Laporan Studi Independen Bersertifikat Kampus Merdeka

Surabaya, 02 Januari 2023

Pembimbing Studi Independen Departemen Teknologi Informasi,  
Institut Teknologi Sepuluh Nopember



Dr. techn. Ir. Raden Venantius Hari Ginardi, M.Sc

NIP: 196505181992031003


**Lembar Pengesahan**  
**DATA SCIENTIST**  
**Di MyEduSolve - PT Dwi Inti Putra**

oleh:  
Najwa Amelia Qorry 'Aina / 5027201001

disetujui dan disahkan sebagai  
Laporan Studi Independen Bersertifikat Kampus Merdeka

Jakarta, 02 Januari 2023

Curriculum Developer MyEduSolve



Tirsa Aulia Puspitasari

## **Abstraksi**

Tujuan dari penulisan laporan akhir Studi Independen ini untuk memberikan laporan atau informasi kepada pihak Kampus Merdeka dan Universitas terkait dengan pelaksanaan kegiatan Studi Independen di MyEduSolve dengan program yang diikuti yaitu #FutureReady: Future-Skills and Career-Ready with MyEduSolve Pathway - Data Scientist dan materi atau kurikulumnya berupa: IT Specialist: Database (SQL), Introduction to Power BI, IT Specialist: Python, Fundamentals of Statistics, Microsoft Azure AI Fundamentals (AI – 900). Kegiatan Studi Independen ini berlangsung selama 18 Agustus – 31 Desember 2022. Hasil dari studi independen dari program Data Scientist adalah dapat memvalidasi kemampuan setiap individu dalam mengoperasikan perangkat lunak seputar pengolahan dan visualisasi data dan juga sertifikasi yang sudah didapat mampu menetapkan standar yang lebih tinggi dan menambah daya saing di industry khususnya pada bidang Data Science.

*Kata Kunci: Data Scientist, Visualisasi, Dataset, Studi Independen, Sertifikasi.*

## **Kata Pengantar**

Puji dan syukur penulis panjatkan ke hadirat Allah SWT., atas karunia-Nya sehingga penulis dapat menyusun Laporan Akhir Studi Independen Bersertifikat di PT Dwi Inti Putra - MyEduSolve 2022. Laporan Akhir Studi Independen Bersertifikat ini merupakan bentuk pertanggungjawaban tertulis atas terlaksananya kegiatan dengan durasi kegiatan kurang lebih 4 bulan yang dimulai dari tanggal 18 Agustus 2022 hingga tanggal 31 Desember 2022.

Pada kesempatan yang baik ini, dengan segenap rasa hormat dan kerendahan hati, penulis mengucapkan rasa terima kasih kepada orang tua yang selalu mendukung, memberikan semangat dan motivasi bagi penulis. Tidak lupa juga saya ucapkan terima kasih kepada :

1. Dosen Wali Dr. techn. Ir. Raden Venantius Hari Ginardi, M.Sc yang telah menjadi pembimbing program Studi Independen Bersertifikat ini.
2. Mentor Pathway, Kak Intan Septiana Sari, yang telah membantu memandu berjalannya kelas Data Science dan proses kegiatan belajar.
3. Mentor Career, Kak Resty El Kholiqy, yang telah mengarahkan dan membimbing dalam kelas Career Mentoring.
4. Semua pihak yang telah membantu penulis dalam pelaksanaan program Studi Independen Bersertifikat.

Penulis menyadari terdapat banyak kekurangan dalam penyusunan laporan akhir ini dan masih jauh dari sempurna. Oleh karena itu penulis mengharapkan saran dan kritik yang bersifat membangun dari semua pihak untuk perbaikan laporan akhir ini. Akhirnya penulis mengharapkan semoga laporan akhir ini dapat memberikan manfaat kepada semua pihak.

Ponorogo, 02 Januari 2023



Najwa Amelia Qorry 'Aina

## Daftar Isi

|  |    |
|--|----|
| Lembar Pengesahan Departemen Teknologi Informasi .....               | 1  |
| Lembar Pengesahan .....  | 2  |
| Abstraksi .....  | 3  |
| Kata Pengantar .....   | 4  |
| Bab I   Pendahuluan .....  | 6  |
| I.1   Latar Belakang .....   | 6  |
| I.2   Lingkup .....  | 7  |
| I.3   Tujuan.....  | 7  |
| Bab II   MyEduSolve – PT Dwi Inti Putra .....                        | 8  |
| II.1   Struktur Organisasi.....                                      | 8  |
| II.2   Lingkup Pekerjaan.....  | 9  |
| II.3   Deskripsi Pekerjaan .....                                     | 9  |
| II.4   Jadwal Kerja .....  | 10 |
| Bab III   Pathway Data Scientist .....                               | 11 |
| III.1   Deskripsi Pembelajaran Pathway Data Scientist .....          | 11 |
| III.2   Proses Pelaksanaan Pembelajaran Pathway Data Scientist ..... | 11 |
| III.3   Pencapaian Hasil Pembelajaran Pathway Data Scientist.....    | 13 |
| Bab IV   Penutup .....   | 15 |
| IV.1   Kesimpulan .....  | 15 |
| IV.2   Saran .....   | 15 |
| Lampiran A – TOR .....   | 16 |
| Lampiran B – Log Activity .....                                      | 19 |
| Lampiran C – Dokumen Teknik .....                                    | 35 |

## **Bab I Pendahuluan**

### **I.1 Latar Belakang**

Sebagai bentuk nyata menunjang aspek keahlian profesional, Kemendikbudristek menyediakan sarana dan prasarana untuk tidak hanya belajar di dalam lingkup kampus saja tetapi dapat mengikuti berbagai macam program yang telah disediakan pada program Kampus Merdeka 2022. Karena menurut Nadiem Makarim selaku Menteri Pendidikan dan Kebudayaan Riset dan Teknologi dalam dunia kerja nantinya dibutuhkan keterpaduan antara pengetahuan akan teori yang telah didapatkan dari bangku perkuliahan dan pelatihan praktik di lapang guna memberikan gambaran tentang dunia kerja yang sebenarnya. Maka dari itu terciptalah salah satu program Studi Independen Kampus Merdeka, salah satu mitra yang bergabung dengan program Kampus Merdeka ini adalah PT. Dwi Inti Putra - MyEduSolve.

PT. Dwi Inti Putra merupakan perusahaan yang bergerak di bidang perdagangan peralatan kantor dan bank modern sejak tahun 2011. Secara konsisten, PT. Dwi Inti Putra terus berkomitmen memberikan pelayanan dan kinerja yang optimal kepada klien. Dalam perkembangannya, PT Dwi Inti Putra bekerja sama dengan Microsoft International untuk melaksanakan program pelatihan dan sertifikasi program-program Microsoft. Sebagai Certiport Authorized Testing Center (CATC), juga menjalin kerja sama dengan lembaga pendidikan tingkat SMA/SMK dan Universitas untuk meningkatkan kecakapan mereka dalam program Microsoft yang berguna untuk kesiapan terjun ke dunia kerja.

MyEduSolve memiliki visi untuk membantu generasi muda Indonesia agar siap berkarier melalui pelatihan literasi digital serta sertifikasi internasional dan misinya adalah mendukung serta memberdayakan generasi muda agar bisa unggul di profesi pilihan mereka serta mampu mengeksekusi ide dan kreativitas mereka dengan baik. Melalui program Studi Independen Kampus Merdeka, MyEduSolve ingin membangun Indonesia yang lebih baik dimana generasi muda dapat lebih unggul dan berdaya.

## **I.2 Lingkup**

Pokok pembelajaran pada program #FutureReady: Future-Skills and Career-Ready with MyEduSolve Pathway - Data Scientist meliputi:

- Fundamental of Statistics
- IT Specialist: Database (SQL)
- IT Specialist: Python
- Introduction to Power BI
- Microsoft Azure AI Fundamentals (AI-900)
- Career Mentoring
- Capstone Project

Dengan rangkaian kegiatan program yang dilaksanakan:

- Pembelajaran Pathway
- Ujian Sertifikasi
- Career Mentoring
- Master Class
- Capstone Project

## **I.3 Tujuan**

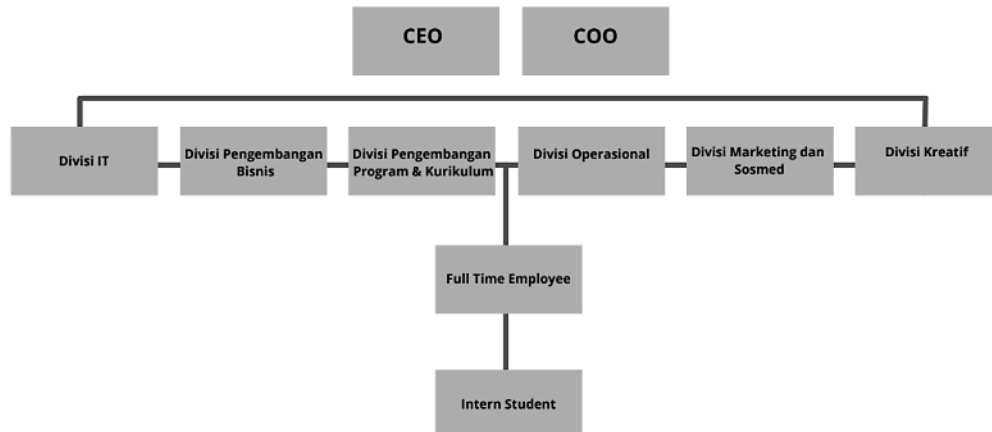
Tujuan pembelajaran ini adalah sebagai mahasiswa yang sedang menyiapkan diri untuk memasuki dunia kerja dan mengembangkan potensinya agar dapat menjadi seorang fresh graduate yang memiliki value di mata perusahaan, mengikuti program Studi Independen di MyEduSolve menjadi salah satu jalan untuk meningkatkan kesempatan dalam memiliki pekerjaan impian. Dengan literasi digital yang cukup, generasi milenial mampu mencari peluang pekerjaan yang lebih banyak. Dengan mengikuti program Pathway Data Scientist, diharapkan penulis mendapatkan pengetahuan dan dapat menguasai terkait program yang diambil, yaitu Data Science.



## Bab II MyEduSolve – PT Dwi Inti Putra

### II.1 Struktur Organisasi

Berikut merupakan struktur organisasi PT. Dwi Inti Putra - MyEduSolve yang memiliki beberapa divisi, meliputi:



- Divisi IT Head: Memperbaiki sistem yang bermasalah ketika user menggunakannya dan melakukan update setiap kali versi terbaru hadir dari suatu sistem maupun aplikasi.
- Divisi Pengembangan Bisnis: Mencari peluang pelanggan baru, menjaga hubungan dengan pelanggan, menyusun dan mempresentasikan rencana pengembangan bisnis dan riset terkait perkembangan pasar dan kompetitor.
- Divisi Pengembangan Program & Kurikulum: Menyusun program dan kegiatan di bidang-bidang yang berkaitan dengan pembelajaran dan memberikan sosialisasi serta pengarahan tentang program dan kurikulum kepada para staff perusahaan.
- Divisi Operasional: Mengurus terkait dengan segala kegiatan perusahaan yang berkait dengan opsional perusahaan
- Divisi Marketing & Sosmed: Mempromosikan program yang dimiliki perusahaan, serta aktif untuk memberikan informasi di media social.
- Divisi Kreatif: Memiliki tugas untuk memberikan ide-ide yang menarik terhadap suatu program atau kegiatan perusahaan serta untuk menarik perhatian yang akan memikat para target audiens.

## **II.2 Lingkup Pekerjaan**

Pada program MSIB di MyEduSolve terdapat serangkaian kegiatan yaitu program Pembelajaran Pathway dimana penulis ditempatkan di kelas Data Scientist D. Selain itu juga terdapat program Career Mentoring dan Master Class. Kemudian di bulan terakhir program terdapat Capstone Project yang dikerjakan secara berkelompok.

## **II.3 Deskripsi Pekerjaan**

Kegiatan studi independen MyEduSolve dibimbing oleh pembimbing kelas pathway dan pembimbing mentoring. Peran pembimbing kelas pathway dalam studi independen ini adalah sebagai fasilitator yang memberikan atau menyampaikan materi serta pembelajaran bagi peserta studi independen sesuai dengan pathway yang telah diambil selama kegiatan studi independen yaitu Data Scientist. Sedangkan peran pembimbing mentoring sebagai fasilitator dalam bidang career untuk memantapkan tujuan karir setelah lulus kuliah sesuai dengan prosedur atau materi yang sudah ditetapkan. Pelaksanaan pembelajaran pada kegiatan studi independen ini meliputi beberapa kelas sebagai berikut:

- **Pembelajaran Pathway Data Scientist**  
Kegiatan dilaksanakan secara daring menggunakan zoom meeting dibimbing oleh mentor Data Science dan dilaksanakan kurang lebih selama empat bulan termasuk dengan pengambilan ujian sertifikasi di certiport. Selama pembelajaran, selain secara teori juga mengerjakan latihan soal dan juga praktik secara langsung dengan menggunakan software data scientist untuk mengetahui sudah sejauh mana pengetahuan kita terhadap materi tersebut.
- **Career Mentoring dan Master Class**  
Kegiatannya meliputi kelas kecil rutin setiap minggu yang dibimbing oleh Career Mentor membahas terkait persiapan rencana karir mulai dari soft skill hingga dokumen berkas seperti CV dan Portofolio. Selain itu juga terdapat sesi 1-on-1 career mentoring, support group, serta masterclass yang diadakan secara berkala untuk menunjang persiapan karir peserta juga.

- Capstone Project

Pada bulan terakhir program terdapat Capstone Project dimana peserta akan mengerjakan project akhir terkait penyelesaian dengan Data Science secara berkelompok yang akan diberikan langsung oleh client dari berbagai organisasi dan perusahaan di Indonesia, dalam hal ini penulis mengerjakan proyek dari perusahaan Difalink.

#### II.4 Jadwal Kerja

Berikut ini adalah jadwal kegiatan yang harus diikuti oleh para peserta:

| Jenis Program                 | Waktu Pelaksanaan                     |
|-------------------------------|---------------------------------------|
| Kelas Data Scientist          | Senin – Jumat<br>19.00-21.00          |
| 1on1 Mentoring                | Satu kali seminggu<br>Durasi 30 menit |
| Group Mentoring / kelas kecil | Setiap Senin<br>13.00-14.30           |
| Support Group                 | Satu kali seminggu<br>Durasi 60 menit |
| Masterclass / kelas besar     | Setiap Rabu<br>13.00-14.30            |
| Social Mixer                  | Satu kali sebulan<br>Durasi 60 menit  |
| Capstone Project              | 1 bulan<br>November - Desember        |

## **Bab III      Pathway Data Scientist**

### **III.1   Deskripsi Pembelajaran Pathway Data Scientist**

Kegiatan pembelajaran Pathway Data Scientist dilakukan dengan durasi 190 jam yaitu waktu pembelajaran 2 jam sehari dalam lima hari selama satu minggu. Pembelajaran dilaksanakan satu persatu mempelajari setiap modul dari kurikulum yang ada yaitu IT Specialist: Database (SQL), IT Specialist: Python, Python for Data Science (Panda, Numpy, Seaborn, Matplotlib), Fundamental of Statistics, dan Machine Learning Microsoft Azure AI Fundamentals (AI-900). Selain itu juga dilaksanakan ujian sertifikasi internasional untuk modul Database dan Python.

Pada kegiatan pembelajaran Career Mentoring berdurasi 23 jam 40 menit dengan waktu pembelajaran 1 jam 30 menit 1 kali pertemuan dalam seminggu. Career mentoring dilaksanakan untuk membantu dalam meningkatkan pemahaman diri, mempelajari softskill yang dibutuhkan dalam dunia kerja abad ke-21: Teamwork, Adaptability, Critical Thinking, Curiosity, Creativity, dll. Dalam program ini, penulis juga memiliki sesi career mentoring 1 o-n-1 yang mempelajari bagaimana menyiapkan mental untuk bersiap menjadi seorang pekerja, mempersiapkan CV, membuat profil LinkedIn, dan menghadapi interview kerja.

Di akhir pembelajaran, terdapat Capstone Project yang merupakan proyek akhir dari pembelajaran Pathway Data Scientist dimana peserta diminta untuk menyelesaikan persoalan dengan Data Science secara berkelompok yang diberikan oleh mitra perusahaan atau organisasi yang bekerja sama, dalam hal ini penulis mengerjakan proyek dari perusahaan Difalink.

### **III.2   Proses Pelaksanaan Pembelajaran Pathway Data Scientist**

Proses pelaksanaan pembelajaran kelas Data Scientist yang dilakukan selama kegiatan Studi Independen di MyEduSolve adalah seperti tahapan berikut:

- Penyampaian materi setiap harinya oleh mentor pathway Data Scientist terkait dengan modul pada minggu tersebut.
- Hands-on atau praktik langsung pada perangkat lunak yang digunakan seperti Database menggunakan Dbeaver, Power BI menggunakan Microsoft

Power BI, dan Python dan Machine Learning menggunakan Jupyter Notebook.

- Tugas pemecahan masalah data scientist dengan small group yang berisi 4-5 orang atau latihan soal sebagai simulasi ujian sertifikasi.
- Ujian sertifikasi dimana untuk lulus sertifikasi score akhir, yaitu minimal 700 dan maksimal 1000. Sertifikasi dilakukan melalui web certiport dan setelah selesai sertifikasi pun sertifikat akan tercantum di halaman My Transcript di web Certiport.

Adapun modul pembelajaran dan waktu pelaksanaannya adalah sebagai berikut:

- Minggu ke 1-4: IT Specialist Database (SQL)
- Minggu ke 5: Power BI
- Minggu ke 6-9: IT Specialist Python
- Minggu ke 10: Fundamental Statistics
- Minggu ke 11-14: Azure AI (Machine Learning)

Pada kelas Career Mentoring sendiri dengan rangkaian kegiatan Group Mentoring atau kelas kecil setiap Senin, Master Class, dan sesi 1-on-1 atau Support Group keseluruhannya dilaksanakan secara daring melalui zoom meeting dengan bahasan di minggu pertama terkait Self Awareness Journal dan tujuan karir yang dimiliki, minggu kedua membahas Goal Setting Journal, minggu ketiga tentang Mindset dan Resiliensi, minggu ke-empat terkait Komunikasi dan Negosiasi, minggu ke-lima membahas CV, Cover Letter, Portofolio, dan Self Saboteur, minggu ke-enam tentang Berpikir Kritis dan Work Value, minggu ke-tujuh membahas LinkedIn dan Job Portal, minggu ke-delapan membahas Berpikir Kreatif, minggu ke-semilan tentang Interview dan FGD, minggu ke-sepuluh tentang Problem Solving dan Decision Making serta Time Management, minggu ke-sebelas tentang Mitos Dunia Kerja, minggu ke-dua belas tentang Teamwork dan simulasi interview, minggu ke-tiga belas tentang Adaptability dan Diversity, kemudian di minggu ke-empat belas yang merupakan minggu formal terakhir adalah perpisahan dan review keseluruhan,

Selanjutnya pada rangkaian kegiatan Capstone Project yang merupakan projek dari klien yang telah bekerja sama dengan MyEduSolve dimana dalam hal ini penulis

mengerjakan proyek dari perusahaan Difalink, yang merupakan social enterprise yang berfokus pada pemberdayaan penyandang disabilitas secara ekonomi melalui pekerjaan, training dan capacity building. Dalam pengerjaannya dibentuk grup yang beranggotakan 4 orang untuk menyelesaikan tugas proyek yaitu mengolah data yang ada di dalam Difalink atau berpotensi dijangkau oleh Difalink secara lebih profesional dan terstruktur. Selama pengerjaan proyek terdapat sesi mentoring bersama dengan pihak Difalink dan sesi mentoring bersama mentor Data Scientist dari MyEduSolve yang dilaksanakan sekali setiap minggunya. Adapun secara berkelompok juga rutin melakukan group meeting untuk membahas proyeknya dengan metode pengerjaan proyek sebagai berikut:

- *Define Problem Statement*: mendefinisikan persoalan dan objektif yang diminta dari Difalink.
- *Data Collecting*: mengumpulkan dataset baik dari Difalink langsung maupun dataset tambahan sebagai bahan analisis dari sumber lain.
- *Data Cleaning*: merapikan dan membersihkan dataset menggunakan google colab dengan python.
- *Data Exploration & Analysis*: mencari insight dari dataset menggunakan google colab dengan python kemudian dianalisis menyesuaikan objektif Difalink.
- *Research*: melakukan riset dari berbagai sumber lain untuk dapat memberikan saran dan rekomendasi kepada Difalink berdasarkan data lapangan lain.
- *Presenting The Result*: menyusun story deck pada power point dari hasil pengerjaan yang kemudian dipresentasikan pada pihak Difalink.

### **III.3 Pencapaian Hasil Pembelajaran Pathway Data Scientist**

Setelah melaksanakan kegiatan Studi Independen di PT. Dwi Inti Putra – MyEduSolve selama kurang lebih lima bulan, penulis berhasil mencapai hasil pembelajaran dengan mengimplementasikannya pada tugas atau proyek yang diberikan, seperti pada setiap rangkaian kegiatan menghasilkan pencapaian pembelajaran berupa:

- Pembelajaran Pathway Data Scientist  
Sertifikasi Internasional IT Specialist: Database, Sertifikasi Internasional IT Specialist: Python, Dashboard Power BI, Machine Learning Project.
- Career Mentoring dan Master Class  
CV ATS Profesional, Cover Letter, Portofolio, LinkedIn Profil Profesional, Time Management, Career Planner, Goal Setting Journal.
- Capstone Project  
Project berupa coding di Google Colab, Dashboard Google Data Studio, dan Power Point Presentation.

## **Bab IV        Penutup**

### **IV.1    Kesimpulan**

Berdasarkan kegiatan Studi Independen yang dilaksanakan penulis di PT. Dwi Inti Putra - MyEduSolve, penulis telah mencapai tujuan yang telah dirumuskan yaitu penulis telah melakukan berbagai kegiatan pembelajaran sesuai dengan yang diinstruksikan oleh mentor pathway dan mentor career, selain itu dalam pelaksanaan kegiatan pembelajaran penulis telah berhasil mengikuti dan lulus ujian sertifikasi program Data Scientist sesuai dengan program yang penulis ambil dalam Studi Independen kampus merdeka ini yaitu Sertifikasi IT Specialist: Database dan IT Specialist: Python. Penulis juga mendapatkan pengetahuan baru selama kegiatan career mentoring, yaitu terkait bagaimana menghadapi dunia kerja yang sesungguhnya nanti, wawasan dan keterampilan karir yang nantinya dapat bermanfaat dalam dunia kerja. Pencapaian tersebut sebagian besar tidak didapatkan di perkuliahan dimana di dunia kerja lebih banyak praktiknya.

### **IV.2    Saran**

Adapun saran untuk Studi Independen selanjutnya yang diajukan penulis dari kegiatan pembelajaran Studi Independen Bersertifikat Kampus Merdeka di PT. Dwi Inti Putra - MyEduSolve periode ini adalah sebagai berikut:

- Akan lebih baik jika ditambahkan program English Skill sehingga dapat membantu pada saat akan menghadapi ujian sertifikasi internasional maupun untuk memasuki dunia kerja.
- Selama program Career Mentoring pembahasan langsung pada dunia kerja seputar konsultasi CV dan LinkedIn serta simulasi interview akan lebih mematangkan kesiapan peserta untuk mempersiapkan memasuki dunia kerja.



**Lampiran A – TOR**  
**TERM OF REFERENCE (TOR)**  
**STUDI INDEPENDEN BERSERTIFIKAT**  
**DATA SCIENTIST DI MYEDUSOLVE**

**A. Rincian Program**

Pathway Data Scientist mempelajari materi mengenai pemahaman dan keterampilan dalam penyelesaian masalah menggunakan pendekatan berbasis data, model prediksi, serta penggunaan Artificial Intelligence dan Machine Learning. Tak hanya itu, pada program ini juga mempersiapkan pesertanya sehingga nantinya siap dalam hal karir juga. Pokok pembelajaran meliputi:

- Fundamental of Statistics
- IT Specialist: Database (SQL)
- IT Specialist: Python
- Introduction to Power BI
- Microsoft Azure AI Fundamentals (AI-900)
- Career Mentoring
- Capstone Project

**B. Benefit Program**

Benefit atau keuntungan yang bisa didapatkan oleh peserta meliputi:

- Pembelajaran intensif bersama mentor profesional dari
- berbagai industri
- Bimbingan karir bersama konsultan karier profesional
- 1 voucher ujian sertifikasi
- 1 voucher retake ujian sertifikasi
- Akses pembelajaran pada GMetrix untuk persiapan ujian
- Akses webinar wawasan industri mengenai pengembangan dan keterampilan diri (masterclass)
- Kuota internet setiap bulan sebesar Rp. 200.000

### C. Jadwal Kegiatan

Berikut ini adalah jadwal kegiatan yang harus diikuti oleh para peserta:

| Jenis Program                 | Waktu Pelaksanaan                     |
|-------------------------------|---------------------------------------|
| Kelas Data Scientist          | Senin – Jumat<br>19.00-21.00          |
| 1on1 Mentoring                | Satu kali seminggu<br>Durasi 30 menit |
| Group Mentoring / kelas kecil | Setiap Senin<br>13.00-14.30           |
| Support Group                 | Satu kali seminggu<br>Durasi 60 menit |
| Masterclass / kelas besar     | Setiap Rabu<br>13.00-14.30            |
| Social Mixer                  | Satu kali sebulan<br>Durasi 60 menit  |
| Capstone Project              | 1 bulan<br>November - Desember        |

### D. Hak dan Kewajiban Peserta

Hak dan kewajiban seluruh peserta selama mengikuti Studi Independen MyEduSolve adalah sebagai berikut:

- Mengikuti program MyEduSolve x Kampus Merdeka selama 21 minggu (1 semester) tanpa dipungut biaya apapun.
- Mendapatkan bimbingan dan fasilitas selama kegiatan pembelajaran oleh mentor kelas dan mentor karier.
- Mendapatkan fasilitas webinar dan workshop.
- Mendapatkan fasilitas kuota internet setiap bulan selama program.
- Menanyakan informasi terkait dengan pembelajaran pada program.
- Mendapatkan transkrip nilai hasil pembelajaran pada akhir program.
- Mendapatkan SKS konversi sesuai rekomendasi hasil akhir, berdasarkan kesepakatan dengan Perguruan Tinggi pada awal program.

- Mendapatkan sertifikat penyelesaian program (bagi peserta yang lulus) dan surat keikutsertaan (bagi peserta tidak lulus) pada akhir program.
- Mengikuti program dengan sungguh-sungguh selama periode berlangsung.
- Mematuhi seluruh peraturan yang berlaku.
- Tidak memberikan akses pembelajaran maupun ujian sertifikasi kepada pihak lain yang tidak terdaftar sebagai peserta SIB MyEduSolve.
- Menghadiri seluruh kegiatan pembelajaran seperti synchronous learning, career mentoring, group mentoring, ujian sertifikasi, webinar, workshop, dan kegiatan tambahan lainnya sesuai dengan ketentuan yang telah ditetapkan agar memperoleh hasil yang maksimal.
- Menyelesaikan pembelajaran asynchronous oleh MyEduSolve.
- Aktif dalam menggunakan fasilitas komunikasi yang disediakan, yaitu Whatsapp, Discord, dan Email.
- Mencatat pembelajaran setiap hari, minggu, bulan pada Logbook yang tersedia di website Kampus Merdeka.
- Melaporkan progres dan rincian pembelajaran kepada DIKTI dan dosen pembimbing di Perguruan Tinggi asal, sesuai dengan ketentuan dan mekanisme yang berlaku.
- Apabila mahasiswa berhalangan hadir dikarenakan sakit atau kondisi darurat lainnya, maka wajib melaporkan hal tersebut kepada mentor.

Demikian Term of Reference ini dibuat, sebagai pedoman dengan harapan dapat dilaksanakan sehingga ada kerjasama yang baik bagi peserta dan mitra.

Mengetahui

Curriculum Developer MyEduSolve



Tirsa Aulia Puspita Sari

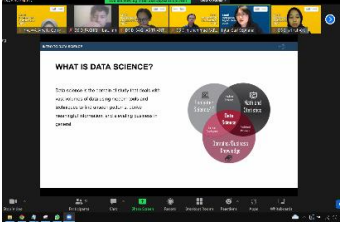
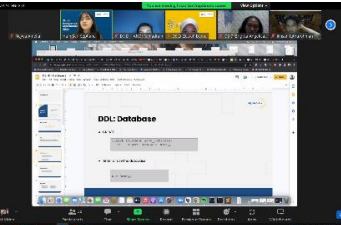
Surabaya, 28 Desember 2022

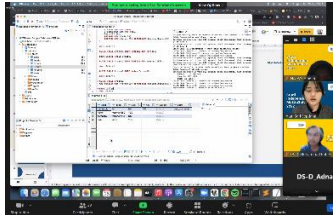
Peserta Program





Najwa Amelia Qorry 'Aina

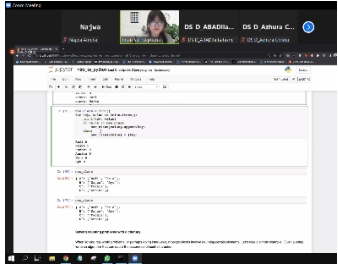
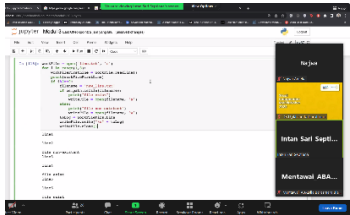
## Lampiran B – Log Activity

| Minggu/Tgl                           | Kegiatan  | Dokumentasi/Hasil  |
|--------------------------------------|---|--|
| Minggu ke-1<br>(18 - 19 Agu<br>2022) | <p>Kegiatan minggu ini Zoom Meeting acara "Pisah Sambut MSIB angkatan 2 dan 3" yang dimulai dengan sambutan dari Mitra, Perguruan Tinggi, Perwakilan Mahasiswa dan Orang tua, serta testimoni terkait MSIB dari perwakilan alumni MSIB angkatan 2. Kemudian terdapat Social Mixer yang merupakan acara perkenalan dan ice breaking sebelum dimulainya kelas dimana seluruh peserta SIB MyEduSolve bertemu dalam Zoom Meeting dan melaksanakan bonding. Kemudian ada pertemuan kelas Data Scientist pertama diisi dengan perkenalan, set rules dan pembagian kelompok tugas, lalu dipaparkan terkait course overview, timeline, assessment, dan weekly assignment, serta yang terakhir pengenalan tools yang akan digunakan selama kegiatan.</p> |   |
| Minggu ke-2<br>(22 - 26 Agu<br>2022) | <p>Pada minggu terdapat kelas pertama Career Mentoring dengan perkenalan dan dilanjutkan dengan materi "Self-Awareness" serta Master Class bertopik Career Building: Generalis vs Specialist. Kemudian pada kelas Data Scientist modul Database membahas tentang:</p>   | <p><a href="#">medium/article/database1</a></p>  <p><a href="#">selfawareness-journal</a></p> |

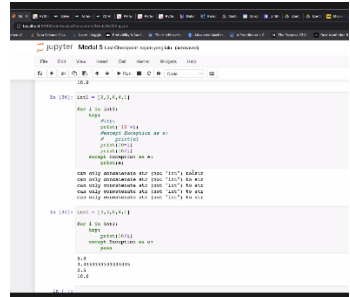
|                                      |  |   |
|--------------------------------------|--|---|
|                                      | <ul style="list-style-type: none"> <li>- Database Schema: berisi table, data types, relationship, dan user roles</li> <li>- Mengorganize data ke tiap entitas dan mendetermines relationship antar entitas serta data typesnya</li> <li>- Relationship: one to one; one to many; many to many</li> <li>- Normalization: 0NF; 1NF; 2NF; 3NF; BCNF; 4NF; 5NF; 6NF</li> <li>- Data Manipulation Language DML, Data Definition Language (DDL), Data Control Language (DCL).</li> </ul> <p>Tak lupa di akhir materi dilaksanakan exercise hands-on dalam melakukan normalization database dan praktik sintaks pada DBeaver.</p> |   |
| Minggu ke-3<br>(29 Agu - 2 Sep 2022) | <p>Kegiatan minggu pada kelas Career Mentoring membahas Ikigai dan Goal Setting. Adapun di kelas malam pathway Data Scientist modul Database membahas topik utama DML dan Query Data Retrieval:</p> <p>Pertama, dalam DML fokus pada command insert, selanjutnya mengerjakan hands-on exercise pada DBeaver bersama kelompok untuk mempraktikkannya.</p> <p>Di pertemuan berikutnya melakukan insert dengan diberikan sample data set csv yang perlu di alter tipe data dan Primary Key Foreign Key-nya sehingga sample data set</p>   | <p><a href="https://medium.com/article/database2">medium/article/database2</a></p>  <p><a href="#">goalsetting-journal</a></p> |


|   |  |  |
|---|--|--|
|   | <p>tersebut berelasi satu sama lain dan memiliki tipe data yang sesuai sehingga dapat di import ke dalam database pada DBeaver. Selanjutnya dibahas query terkait Retrieval Data, CTE, Inner Join, Outer Join, Left Join, Right Join, dan Cross Join dan mencoba hands-on pada data set yang ada dengan menggunakan berbagai command seperti Join, Operator (sum, avg, count, max). Dilanjutkan dengan mengerjakan exercise kelompok dengan studi case yang berbagai macam lalu dibahas bersama di kelas.</p>  |  |
| <p>Minggu ke-4<br/>(5 - 9 Sep 2022)</p> | <p>Minggu ini dimulai di kelas career mentoring membahas tentang Mindset dan Resiliansi. Kemudian di kelas Data Scientist modul Database membahas:</p> <ul style="list-style-type: none"> <li>- Views dan Procedure dengan command utama create, select, call, dan drop</li> <li>- Index: objek dalam database yang mempercepat proses pencarian (query) data</li> <li>- Index type: Btree; Rtree; Fulltext; Hash</li> <li>- Permission: memberikan hak akses tertentu kepada pengguna oleh database administrator</li> <li>- Permission type: create, select, insert, update, delete, drop</li> <li>- Wildcard: menspesifikkan pattern</li> <li>- Regular Expression: biasa digunakan untuk memvalidasi suatu string</li> </ul> | <p><a href="#">certification/database</a></p>  <p>INFORMATION TECHNOLOGY SPECIALIST</p> <p>Najwa Amelia Qorry 'Aina</p> <p>has successfully completed the certification requirements for Databases</p> <p>September 15, 2022</p> <p>Dr. Gory A. Gates<br/>President &amp; CEO<br/>Pearson Education</p> <p>Pivotal</p> <p>CERTIPORT CERTNEXUS Pearson</p> |

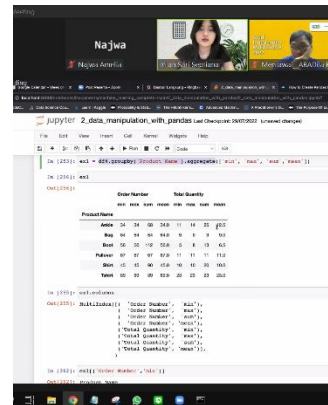
|   |   |   |
|---|---|---|
|   | <p>Kemudian dilanjutkan dengan praktik pada DBeaver dengan case data langsung.</p> <p>Selain itu juga membahas terkait persiapan ujian sertifikasi dengan latihan soal.</p>   |   |
| <p>Minggu ke-5<br/>(12 - 16 Sep 2022)</p> | <p>Agenda minggu ini yaitu ada kelas Career Mentoring yang membahas "Effective Communication" dan sharing session 1on1 yang membahas Goal Setting Journal.</p> <p>Kemudian di kelas Data Scientist mulai memasuki modul selanjutnya: Power BI. Di kelas melakukan instalasi PowerBI bersama kemudian mencoba mengeksplor terkait semua fitur dan kegunaan dari PowerBI untuk menunjang Data Scientist.</p> <p>Selain itu, dibahas juga topik Statistical Data Type:</p> <ul style="list-style-type: none"> <li>- Categorical Qualitative Data: Nominal; Ordinal</li> <li>- Numerical Qualitative Data: Discrete; Continuous</li> </ul> <p>Serta dari suatu dataset dilakukan Data Visualization praktik pada Power BI: Bar Chart; Pie Chart; Line Chart; Scatter Plot; Area Chart; Bubble Chart; Combine Chart.</p> <p>Selanjutnya diberikan suatu sample dataset dari customer bank, kemudian bersama dengan kelompok mengerjakan tugas Data Visualization dengan Power BI untuk mendapatkan insight dan dapat menganalisis karakteristik customer yang meninggalkan</p> | <p><a href="https://canva.com/ppt/powerBI/">canva/ppt/powerBI</a></p>  |


|                                   |  |  |
|-----------------------------------|--|--|
|                                   | bank dan diminta untuk present the data dengan story telling yang sistematis.  |  |
| Minggu ke-6<br>(19 - 23 Sep 2022) | <p>Pada kelas Karir Mentoring membahas terkait "Creative Outstanding CV &amp; Portofolio".</p> <p>Di kelas Data Scientist memasuki modul Python sehingga menginstall Python dan Jupyter Notebook untuk keperluan selama pembelajaran. Kemudian mempelajari:</p> <ul style="list-style-type: none"> <li>- Unit 1 pada handbook modul Python yang diberikan dimana menyangkut: Tipe data; Operasi; Data operator; Struktur data.</li> <li>- Unit 2 yaitu terkait "Flow Control with Decisions and Loops" yang mana mencakup: <ul style="list-style-type: none"> <li>- Branching (if, elif, else, nested if)</li> <li>- Iteration (while, for, break, continue, pass, nested loops)</li> </ul> </li> </ul> <p>Selanjutnya dilakukan bersama latihan dengan live coding pada Jupyter Notebook untuk mempraktikkan setiap materi.</p> | <p><a href="#">jupyter/code/python</a></p>  <p><a href="#">cv-coverletter</a></p> |
| Minggu ke-7<br>(26 - 30 Sep 2022) | <p>Agenda minggu ini di kelas Data Scientist modul Python dimana kali ini membahas lebih dalam terkait:</p> <ul style="list-style-type: none"> <li>- Struktur Data: List, Set, Tuple.</li> <li>- Function: Lambda function, Built in function, Map function, Filter function, List comprehension, Enumerate function, Zip function.</li> </ul>   | <p><a href="#">jupyter/code/python</a></p>                                      |



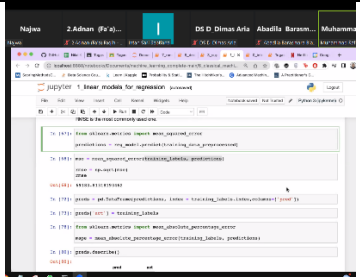
|   |   |  |
|---|---|--|
|   | <ul style="list-style-type: none"> <li>- Input and Output Operations: Open and close, Read and write, Append, With statement, Check existences.</li> <li>- Code Documentation and Structure</li> </ul> <p>Tidak lupa juga disertai dengan praktik langsung dalam jupyter notebook dengan berbagai case.</p> <p>Pembelajaran dilaksanakan secara live coding mempraktikkan case pada jupyter notebook dan diberikan banyak exercise serta test case yang berkaitan dengan struktur data dari python.</p> <p>Pada kelas Career Mentoring membahas topik "Build Critical Thinking for Gen Z" dan sesi sharing 1on1 dengan mentor dari kelas Karir Mentoring yang membahas terkait hasil dari tes "Self-Saboteur", serta Kelas Besar Career Mentoring membahas terkait topik berikut "Remarkable Power of Small Habit &amp; Time Management".</p> |  |
| <p>Minggu ke-8<br/>(3 - 7 Okt 2022)</p> | <p>Pada kelas Data Scientist modul Python minggu ini mempelajari "Troubleshooting and Error Handling" yang mencakup:</p> <ul style="list-style-type: none"> <li>- Error classes: syntax errors, logic errors, runtime errors.</li> <li>- Handle exceptions: try block, except block, else block, finally block, raise an exception.</li> </ul> <p>Kemudian materi "Operations using Modules and Tools" yang mana bahasannya</p>   | <p><a href="https://jupyter.org/code/python">jupyter/code/python</a></p>  |

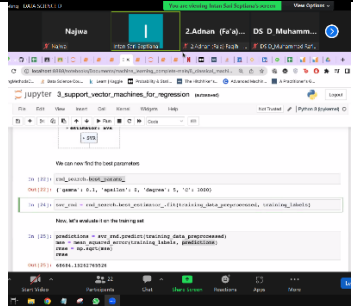

|   |   |   |
|---|---|---|
|   | <p>mencakup: Math module, Sys dan IO module, OS dan OS.path module, Random module, Datetime module.</p> <p>Selanjutnya memperdalam logic terkait manipulasi List dan String dalam python melalui exercise yang dibahas bersama secara live coding.</p> <p>Pada kelas Career Mentoring dengan topik "The Power of LinkedIn for Career" dan sesi sharing 1on1 mentoring yang membahas dan mereview konten dari CV ATS, Cover Letter, dan Portofolio.</p>  |   |
| <p>Minggu ke-9<br/>(10 - 14 Okt 2022)</p> | <p>Agenda minggu ini melakukan persiapan ujian sertifikasi dengan latihan soal bersama dan memperbanyak exercise logic dengan live coding. Kemudian juga mempelajari terkait Data Manipulation menggunakan Pandas secara hands-on pada jupyter notebook dengan cakupan materi: Creating Series; Creating DataFrame; Data Indexing; Selection; Iteration.</p> <ul style="list-style-type: none"> <li>- Dealing with missing data: check missing values, remove missing values, fill the missing values</li> </ul> <p>Dan ditutup dengan exercise dengan case tertentu.</p> <ul style="list-style-type: none"> <li>- Operations of Pandas: info, keys, head, tail, values, shape, unique values</li> <li>- Functions in DataFrames: sort, map, applying function</li> </ul> | <p><a href="#">certification/python</a></p>  <p>IT INFORMATION TECHNOLOGY SPECIALIST</p> <p>Najwa Amelia Qorry 'Aina</p> <p>has successfully completed the certification requirements for Python</p> <p>November 18, 2022</p> <p>Dr. Gary A. Gates<br/>President, IACET<br/>Pearson VUE</p> <p>CERTiPORT CERTNEXUS Pearson</p> |

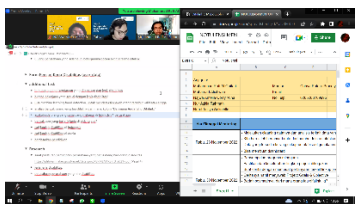
|                                    |  |   |
|------------------------------------|--|---|
|                                    | <ul style="list-style-type: none"> <li>- Aggreation Methods: describe, transpose, mode, max, min, mean, median, std, var, sum, prod</li> <li>- Combining Datasets: Group by, Concatenate, Join, Merge</li> <li>- Working with CSV and Excel in pandas</li> <li>- Exploratory Data Analysis (EDA)</li> </ul> <p>Pada kelas Career Mentoring membahas topik terkait Creative Thinking dan sesi 1on1 Career Mentoring bersama Mentor yang review Linkedin profile, serta di kelas besar yang membahas tentang Personal Branding, Manners, dan Grooming.</p>   |   |
| Minggu ke-10<br>(17 - 21 Okt 2022) | <p>Minggu ini Dalam kelas Data Science mulai memasuki modul baru yaitu Statistics secara teori dan praktik pada coding python numpy yang mencakup:</p> <ul style="list-style-type: none"> <li>- Type of Analytics: descriptive; diagnostic; predictive; prescriptive</li> <li>- Mean, Median, Mode, Percentil, IQR</li> <li>- Skewness: right, left</li> <li>- Kurtosis: Mesokurtic, Leptokurtic, Platykurtic</li> <li>- Probability: Conditional Prob, Independent Events, Mutually Exclusive, Bayes Theorem</li> <li>- Distribution: Normal, Log-Normal, Bernoulli, Binomial, Poisson, Uniform</li> <li>- A/B Testing</li> </ul> <p>Pada kelas Career Mentoring membahas terkait "Successful Job Interview and FGD",</p> | <p><a href="#">pdf/summary/statistics</a></p>  |

|  |   |  |
|--|---|--|
|  | <p>lalu Masterclass Career Mentoring yang bertopik "What to do When Your Career Plans Unrelated with Your Degree", terdapat juga sesi Support Group Discussion dengan melakukan simulasi FGD.</p>   |  |
| <p>Minggu ke-11<br/>(24 - 28 Okt 2022)</p> | <p>Pada kelas Data Science mulai memasuki modul baru yaitu Machine Learning yang dimulai dengan pengenalan dari Machine Learning dan materi utama:</p> <ul style="list-style-type: none"> <li>- Data Preprocessing: Handling missing values; Encode categorical features; Scalling the numeric features</li> <li>- Selecting and training a model:             <ol style="list-style-type: none"> <li>1. the scope of the problem</li> <li>2. the size of the dataset</li> <li>3. the level of interpretability</li> <li>4. training time</li> </ol> </li> <li>- Evaluation Metrics: regression; classification</li> <li>- The challenges of training ML systems: Underfitting (Low Bias); Underfitting (High Bias); Overfitting (High Variance).</li> </ul> <p>Dalam kelas Career Mentoring membahas terkait Problem Solving dan Cara Mengasahnya, lalu sesi sharing 1on1 Mentoring membahas topik Time Management dan Habit Worksheet. Di Kelas Besar Career Mentoring yang membahas sharing session terkait "Successful User Interview".</p> | <p><a href="#">jupyter/code/MLpython</a></p>  <p><a href="#">timemanagement-worksheet</a></p> |

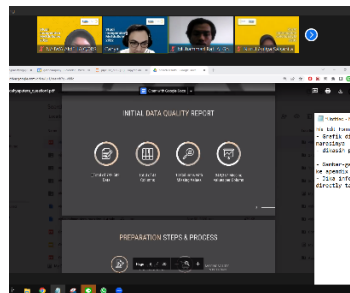
|   |   |  |
|---|---|--|
| <p>Minggu ke-12<br/>(31 Okt - 4 Nov 2022)</p> | <p>Minggu ini Kelas Data Scientist dilaksanakan ujian sertifikasi Python. Kemudian melanjutkan modul Machine Learning dengan hands-on pada jupyter notebook:</p> <ul style="list-style-type: none"> <li>- Exploratory Data Analysis: Visualisasi Seaborn, Relational Plots, Distribution Plots</li> <li>- Encoding Categorical Features: Mapping Method, Ordinary Encoding, Label Encoding, Pandas Dummies, OneHot Encoding</li> <li>- Feature Scalling: Normalization, Standardization, Robust scaling</li> </ul> <p>Minggu ini juga mempresentasi tugas kelompok terkait dengan EDA atau Exploratory Data Analysis dari dataset "taxis" dan menjelaskan insight yang didapat serta visualisasinya pada jupyter notebook.</p> <p>Di kelas Career Mentoring membahas terkait fakta dan mitos yang ada di dunia kerja dan bagaimana cara menyikapinya dengan bijak, terdapat juga sesi Master Class dengan topik "Financial Freedom How to Get There?"</p> | <p><a href="#">jupyter/code/MLpython</a></p>  |
| <p>Minggu ke-13<br/>(7 - 11 Nov 2022)</p>     | <p>Pada kelas Data Science mempelajari Machine Learning - Linear Regression dimana membahas bagaimana regresi menggunakan Linear model. Dijelaskan</p>  | <p><a href="#">jupyter/code/MLpython</a></p>   |

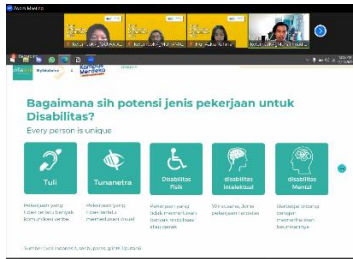
|                                    |   |  |
|------------------------------------|---|--|
|                                    | <p>juga standar langkah dalam pengerjaan Machine Learning Project, yaitu:</p> <ol style="list-style-type: none"> <li>1. Problem formulation</li> <li>2. Collecting the data</li> <li>3. Exploratory Data Analysis</li> <li>4. Data Processing</li> <li>5. Choosing and Training a model</li> <li>6. Evaluating a model</li> <li>7. Improving a model</li> </ol> <p>Selanjutnya membahas mengenai Model Evaluation: MSE, RMSE, MAE</p> <p>Regression pada Machine Learning, yaitu model-model lainnya selain Linear model seperti: Support Vector Machines, Decision Trees, Random Forest</p> <p>Kemudian kami diberikan dataset yang perlu diolah dengan mengikuti langkah pengerjaan tersebut sebagai hands-on practice pada jupyter notebook.</p> <p>Di kelas Career Mentoring yang membahas tentang "Adaptability &amp; Respect for diversity, dealing with difficult people" serta pembekalan Capstone Project, lalu sesi 1-on-1 Career Mentoring kali ini membahas terkait simulasi interview.</p> |  <p><a href="#">careerplanner-journal</a></p> |
| Minggu ke-14<br>(14 - 18 Nov 2022) | <p>Minggu ini pada kelas Data Scientist terdapat materi Machine Learning classification yaitu linear models for classification dan step-step pengerjaan proyeknya seperti:</p>  | <p><a href="#">jupyter/code/MLpython</a></p>   |

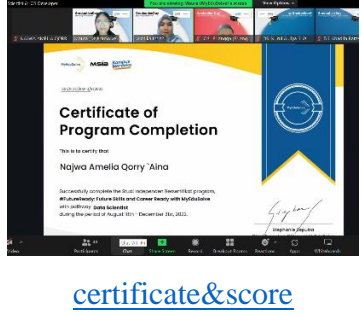

|                                    |   |   |
|------------------------------------|---|---|
|                                    | <ol style="list-style-type: none"> <li>1. Problem formulation</li> <li>2. Finding data</li> <li>3. Exploring insights in data or EDA</li> <li>4. Data preprocessing</li> <li>5. Choosing and training a model</li> <li>6. Evaluating a model</li> </ol> <p>Membahas Classification pada Machine Learning, yaitu model-model lainnya selain Linear model seperti: Support Vector Machines, Decision Trees, Random Forest. Kemudian secara berkelompok mengerjakan dataset yang telah diberikan sebagai practice hands-on dan kemudian presentasi dari hasil tugas kelompok proyek Machine Learning untuk classification dan evaluasi bersama. Pada kelas Career Mentoring kali ini bertopik Career Teamwork, serta sesi 1-on-1 Career Mentoring kali ini membahas terkait evaluasi dua arah dan membahas career planner. Selanjutnya terdapat Kelas Besar dengan topik Salary Negotiation bersama HR Practitioner dan Career Coach. Pada minggu ini juga dilaksanakan sesi onboarding Capstone Project bersama dengan mitra perusahaan Difalink terkait pemaparan proyek akhirnya.</p> |  <p><a href="#">final/CareerMentoring</a></p>    |
| Minggu ke-15<br>(21 - 25 Nov 2022) | Agenda minggu ini adalah sesi briefing Capstone Project yang menjadi tugas akhir pada kegiatan Studi Independen di MyEduSolve. Capstone Project untuk   | <p><a href="#">notion/capstoneProject</a></p>  |

|  |  |   |
|--|--|---|
|  | <p>pathway Data Scientist ini dilaksanakan dengan bekerja sama pada mitra Difalink yang merupakan social enterprise yang berfokus pada pemberdayaan penyandang disabilitas secara ekonomi melalui pekerjaan, training dan capacity building. Dalam pengerjaannya dibentuk grup yang beranggotakan 4 orang untuk menyelesaikan tugas proyek yaitu mengolah data yang ada di dalam Difalink atau berpotensi dijangkau oleh Difalink secara lebih profesional dan terstruktur.</p> <p>Kami melaksanakan meet group team Capstone Project dan membahas secara general terkait prject goals dan objective-nya serta pembagian tugas untuk tiap individu anggota kelompok.</p> <p>Lalu juga terdapat sesi mentoring Capstone Project dengan mentor kelas yang membahas terkait action plan progress dalam penyelesaian proyek.</p> |   |
| Minggu ke-16<br>(28 Nov - 2<br>Des 2022) | <p>Agenda pada hari Senin adalah Weekly meet bersama Mitra kerjasama Capstone Project Data Scientist MyEduSolve yaitu Difalink, dimana dalam meet tersebut membahas progress proyek dari setiap kelompok serta arahan lanjutan terkait data yang disediakan oleh Difalink. Kemudian malamnya melakukan group meeting untuk melanjutkan penugasannya.</p>   | <p><a href="#">notion/capstoneProject</a></p> <p><a href="#">datastudio/dashboard</a></p>  |



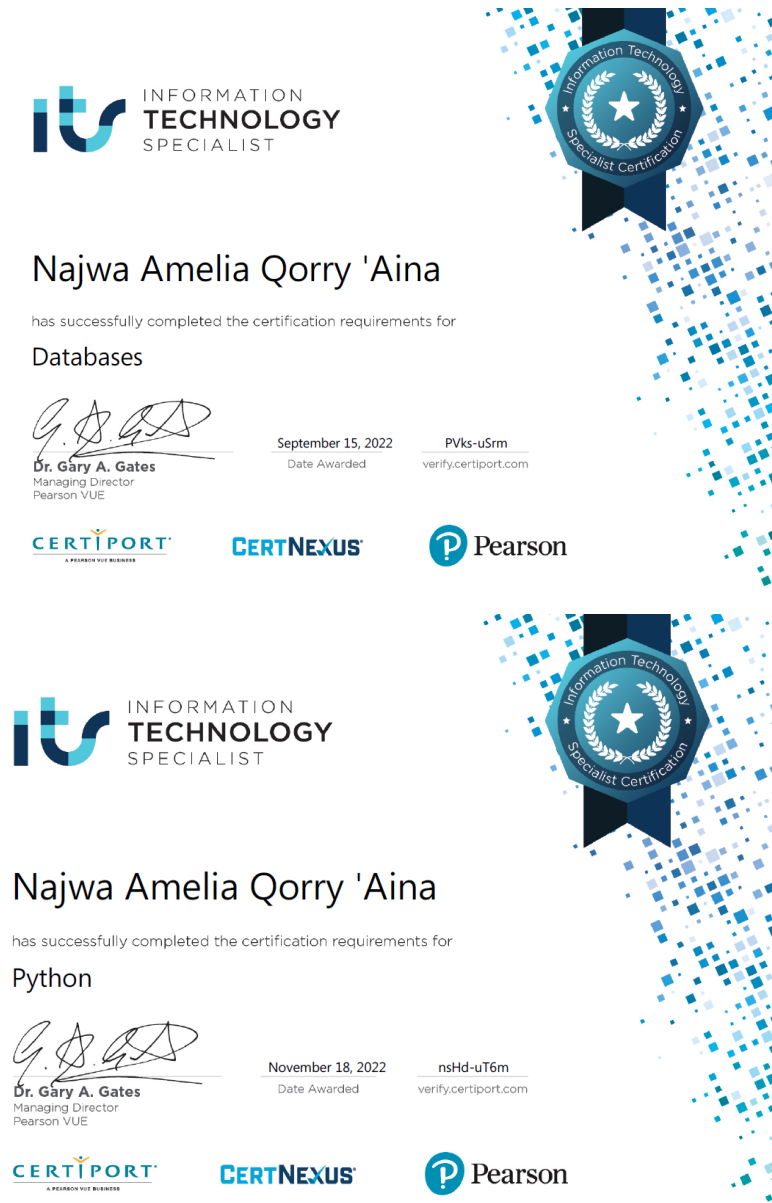
|  |   |   |
|--|---|---|
|  | <p>Selanjutnya kerja kelompok dengan meet group team Capstone Project dimana kami tracking tugas pada Notion dan mengerjakan projectnya atau pengolahan datanya melalui Google Colaborator. Saya yang diamanahi terkait Data Cleaning dan EDA (Exploratory Data Analysis) mengolah data area pencari kerja disabilitas yang aktif mencari pekerjaan dari Difalink dan mengerjakannya pada Python yaitu melakukan visualisasi dan analisis insightnya serta menyeragamkan nama-nama kolom antar data-data yang diberikan oleh Difalink sehingga memudahkan pengolahan datanya kita akan melakukan merging informasi. Pada sesi mentoring progres Capstone Project bersama mentor kelas melakukan penyampaian progress dan pemberian masukan dari mentor, lalu mulai diminta untuk menyusun deck power point dan action item yang perlu Difalink lakukan.</p> |   |
| <p>Minggu ke-17<br/>(5 - 9 Des 2022)</p> | <p>Minggu ini seperti biasa terdapat Weekly meet bersama Mitra kerjasama Capstone Project Data Scientist MyEduSolve yaitu Difalink, dimana dalam meet tersebut membahas progress proyek dari setiap kelompok yaitu:</p> <ul style="list-style-type: none"> <li>- cleaning data yang sudah diterima kerja rekruter</li> </ul>  | <p><a href="https://www.notion.so/capstoneProject">notion/capstoneProject</a></p>  |

|                                    |  |  |
|------------------------------------|--|--|
|                                    | <ul style="list-style-type: none"> <li>- sudah menganalisis beberapa data rekrut tersebut untuk mengetahui berdasarkan area, pendidikan, dan gender</li> <li>- jenis disabilitas diseragamkan dengan form user, area yang digunakan area diterima kerja dan area provinsi tempat tinggal</li> <li>- sudah mulai menyusun deck presentasi power point</li> </ul> <p>Selanjutnya pada sesi mentoring progres Capstone Project bersama mentor kelas melakukan simulasi presentasi dan diberikan beberapa evaluasi.</p>  |  |
| Minggu ke-18<br>(12 - 16 Des 2022) | <p>Kegiatan minggu ini adalah group meeting membahas Capstone Project yaitu fokus dalam hal penyusunan deck atau story berupa power point dari hasil pengerjaan proyek untuk dipresentasikan ke Mitra Difalink serta ke mentor kelas.</p> <p>Kemudian di minggu ini yang merupakan demo day Capstone Project Data Scientist MyEduSolve dengan mitra Difalink dilakukan presentasi hasil tugas kami dalam bentuk deck atau story power point, dashboard, data research, dan google colab (coding) yang kemudian juga dikumpulkan pada platform yang disediakan sebagai nilai dari proyek akhir.</p> | <p><a href="#">canva/ppt/capstoneProject</a></p> <p><a href="#">datastudio/dashboard</a></p> <p><a href="#">googlecolab/code</a></p>  |

|  |   |   |
|--|---|---|
| <p>Minggu ke-19<br/>(19 - 23 Des<br/>2022)</p> | <p>Agenda pada minggu ini adalah Graduation Day yaitu merayakan segala proses yang telah dilalui dengan susunan acara:</p> <ul style="list-style-type: none"> <li>- Graduation Ceremony</li> <li>- Talk Show with Alumni</li> <li>- Student Awards</li> </ul> |  <p><a href="#">certificate&amp;score</a></p>  |
| <p>Minggu ke-20<br/>(26 - 30 Des<br/>2022)</p> | <p>Kegiatan pada minggu ini oleh karena keseluruhan kegiatan formal dari Studi Independen MyEduSolve telah berakhir, sehingga hanya diisi dengan self-learning and exploring serta mulai menyusun laporan akhir untuk kegiatan MSIB ini.</p>                  | <p>LAPORAN AKHIR<br/>MAGANG &amp; STUDI INDEPENDEN BERSERTIFIKAT<br/>DATA SCIENTIST<br/>DI MYEDUSOLVE - PT DWI INTI PUTRA</p> <p>Dijadkan untuk memenuhi persyaratan kelulusan<br/>Program MSIB MBKM</p> <p>oleh:<br/>Najwa Amelia Qorry 'Aina / 5027201001</p>  |

## Lampiran C – Dokumen Teknik

Dokumen sertifikasi internasional IT Specialist: Databases & IT Specialist: Python.



Link dari dokumentasi Dashboard Power BI Project:

- [canva/ppt/powerBI](https://canva.com/ppt/powerBI)

Link dari dokumentasi hasil dari Capstone Project:

- [canva/ppt/capstoneProject](https://canva.com/ppt/capstoneProject)
- [datastudio/dashboard](https://datastudio.google.com/dashboard)

Dokumentasi pengerjaan salah satu proyek dari kelas Data Scientist pada Jupyter Notebook.

## A. Problem Formulation

Dataset Taxis berisi informasi yang dimiliki oleh perusahaan taksi berupa tarif, jarak, penumpang, & penghasilan supir. Sebagai data scientist diminta untuk memprediksi tip yang didapat oleh supir dengan membangun machine learning

## B. Collecting the library and data

```
import numpy as np
import pandas as pd
import seaborn as sns
import plotly.express as px
import matplotlib.pyplot as plt

import warnings
warnings.filterwarnings('ignore')
```

```
taxis = sns.load_dataset('taxis')
```

```
taxis.head()
```

|   | pickup              | dropoff             | passengers | distance | fare | tip  | tolls | total | color  | payment     | pickup_zone           | dropoff_zone          | pickup_borough | dropoff_borough | pickup_year |
|---|---------------------|---------------------|------------|----------|------|------|-------|-------|--------|-------------|-----------------------|-----------------------|----------------|-----------------|-------------|
| 0 | 2019-03-23 20:21:09 | 2019-03-23 20:27:24 | 1          | 1.60     | 7.0  | 2.15 | 0.0   | 12.95 | yellow | credit card | Lenox Hill West       | UN/Turtle Bay South   | Manhattan      | Manhattan       | 201         |
| 1 | 2019-03-04 16:11:55 | 2019-03-04 16:19:00 | 1          | 0.79     | 5.0  | 0.00 | 0.0   | 9.30  | yellow | cash        | Upper West Side South | Upper West Side South | Manhattan      | Manhattan       | 201         |
| 2 | 2019-03-27 17:53:01 | 2019-03-27 18:00:25 | 1          | 1.37     | 7.5  | 2.36 | 0.0   | 14.16 | yellow | credit card | Alphabet City         | West Village          | Manhattan      | Manhattan       | 201         |
| 3 | 2019-03-10 04:00:50 | 2019-03-10 04:05:54 | 1          | 7.70     | 27.0 | 6.15 | 0.0   | 36.95 | yellow | credit card | Hudson Sq             | Yorkville West        | Manhattan      | Manhattan       | 201         |

## C. Exploratory Data Analysis

```
from sklearn.model_selection import train_test_split
```

```
train_data, test_data = train_test_split(taxis, test_size=0.1, random_state=20)
```

```
print('The size of training data is: {} \nThe size of testing data is: {}'.format(len(train_data), len(test_data)))
```

```
The size of training data is: 5789
```

```
The size of testing data is: 644
```

### Checking data statistics

```
train_data.describe(include='all').transpose()
```

|            | count  | unique | top                 | freq | first               | last                | mean      | std       | min | 25%  | 50%   | 75%  | max    |
|------------|--------|--------|---------------------|------|---------------------|---------------------|-----------|-----------|-----|------|-------|------|--------|
| pickup     | 5789   | 5773   | 2019-03-28 08:21:02 | 2    | 2019-02-28 23:29:03 | 2019-03-31 23:43:45 | NaN       | NaN       | NaN | NaN  | NaN   | NaN  | NaN    |
| dropoff    | 5789   | 5782   | 2019-03-01 19:18:55 | 2    | NaT                 | NaT                 | NaN       | NaN       | NaN | NaN  | NaN   | NaN  | NaN    |
| passengers | 5789.0 | NaN    | NaN                 | NaN  | NaT                 | NaT                 | 1.536189  | 1.200068  | 0.0 | 1.0  | 1.0   | 2.0  | 6.0    |
| distance   | 5789.0 | NaN    | NaN                 | NaN  | NaT                 | NaT                 | 3.020888  | 3.820907  | 0.0 | 0.97 | 1.63  | 3.22 | 36.7   |
| fare       | 5789.0 | NaN    | NaN                 | NaN  | NaT                 | NaT                 | 13.079128 | 11.615997 | 1.0 | 6.5  | 9.0   | 15.0 | 150.0  |
| tip        | 5789.0 | NaN    | NaN                 | NaN  | NaT                 | NaT                 | 1.980869  | 2.470192  | 0.0 | 0.0  | 1.66  | 2.76 | 33.2   |
| tolls      | 5789.0 | NaN    | NaN                 | NaN  | NaT                 | NaT                 | 0.314873  | 1.39418   | 0.0 | 0.0  | 0.0   | 0.0  | 24.02  |
| total      | 5789.0 | NaN    | NaN                 | NaN  | NaT                 | NaT                 | 18.503916 | 13.896511 | 1.3 | 10.8 | 14.15 | 20.3 | 174.82 |
| color      | 5789   | 2      | yellow              | 4910 | NaT                 | NaT                 | NaN       | NaN       | NaN | NaN  | NaN   | NaN  | NaN    |
| payment    | 5748   | 2      | credit card         | 4097 | NaT                 | NaT                 | NaN       | NaN       | NaN | NaN  | NaN   | NaN  | NaN    |

### Checking Missing Values

```
train_data.isnull().sum()

pickup      0
dropoff     0
passengers  0
distance    0
fare        0
tip         0
tolls       0
total       0
color       0
payment     41
pickup_zone 23
dropoff_zone 40
pickup_borough 23
dropoff_borough 40
pickup_year 0
pickup_month 0
pickup_hour 0
pickup_date 0
pickup_dow  0
dtype: int64
```

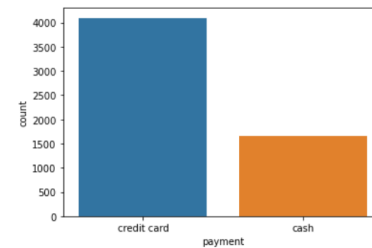
### Checking Values in the Categorical Feature(s)

```
set(train_data.columns) - set(train_data.describe().columns)

{'color', 'payment'}
```

```
sns.countplot(data=train_data, x='payment')
```

<AxesSubplot:xlabel='payment', ylabel='count'>



### Removing Noise Features

```
train_data.drop(columns=['dropoff', 'pickup', 'tolls', 'pickup_year', 'dropoff_borough', 'pickup_borough', 'dropoff_zone', 'pickup_zone'])
```

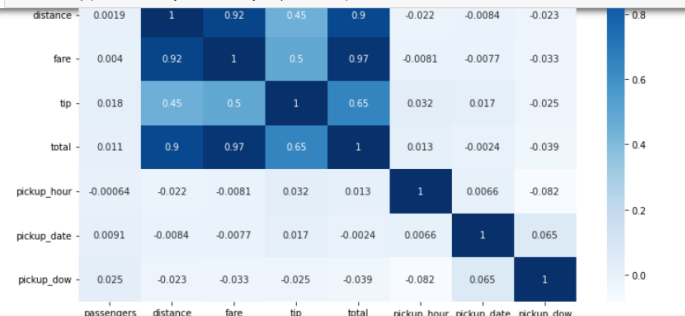
### Checking Correlation Between Features

```
correlation = train_data.corr()
correlation['tip']
```

```
passengers    0.017600
distance      0.454803
fare          0.496169
tip           1.000000
total         0.652495
pickup_hour   0.032075
pickup_date   0.016816
pickup_dow    -0.025106
Name: tip, dtype: float64
```

```
plt.figure(figsize=(12,7))
```

```
sns.heatmap(correlation, annot=True, cmap="Blues")
```



## D. Data Processing

### Handling Missing Values

```
training_input_data.isna().sum()
```

```
passengers    0
distance      0
fare          0
tip           0
total         0
payment       41
pickup_hour   0
pickup_date   0
pickup_dow    0
dtype: int64
```

```
cond = training_input_data['tip'] == 0.0
training_input_data['payment'] = training_input_data['payment'].fillna(cond.map({True:'cash', False:'credit card'}))
```

## Encoding Categorical Features

### Ordinary Encoding

```
from sklearn.preprocessing import OrdinalEncoder

cats_feats = training_input_data[['payment']]

encoder = OrdinalEncoder()

cats_encoded = encoder.fit_transform(cats_feats)
```

```
cats_encoded
```

```
# cash = 0
# credit card = 1
```

```
array([[1.],
       [0.],
       [0.],
       ...,
       [1.],
       [0.],
       [1.]])
```

```
training_input_data[['payment']] = cats_encoded_df
training_input_data.head()
```

|      | passengers | distance | fare | total | payment | pickup_hour | pickup_date | pickup_dow |
|------|------------|----------|------|-------|---------|-------------|-------------|------------|
| 1570 | 1          | 1.20     | 8.0  | 13.55 | 1.0     | 9           | 7           | 3          |
| 1571 | 5          | 0.64     | 5.0  | 9.30  | 0.0     | 16          | 5           | 1          |
| 4518 | 2          | 1.16     | 9.5  | 12.80 | 0.0     | 14          | 6           | 2          |
| 1541 | 2          | 2.90     | 12.5 | 19.55 | 1.0     | 21          | 14          | 3          |
| 5384 | 1          | 0.68     | 5.0  | 8.30  | 0.0     | 12          | 28          | 3          |

## Scaling Numerical features

### Normalization

```
from sklearn.preprocessing import MinMaxScaler
```

```
scaler = MinMaxScaler()
```

```
num_scaled = scaler.fit_transform(num_feats)
num_scaled
```

```
array([[0.16666667, 0.03269755, 0.04697987, 0.07059705, 0.39130435],
       [0.83333333, 0.01743869, 0.02684564, 0.0461042 , 0.69565217],
       [0.33333333, 0.03160763, 0.05704698, 0.06627478, 0.60869565],
       ...,
       [0.33333333, 0.03242507, 0.04362416, 0.0703089 , 0.52173913],
       [0.5       , 0.02588556, 0.02684564, 0.04034117, 0.56521739],
       [0.33333333, 0.03923706, 0.04697987, 0.07411249, 0.86956522]])
```

```
training_data_preprocessed.head()
```

|      | passengers | distance | fare     | total    | pickup_hour | payment |
|------|------------|----------|----------|----------|-------------|---------|
| 1570 | 0.166667   | 0.032698 | 0.046980 | 0.070597 | 0.391304    | 1.0     |
| 1571 | 0.833333   | 0.017439 | 0.026846 | 0.046104 | 0.695652    | 0.0     |
| 4518 | 0.333333   | 0.031608 | 0.057047 | 0.066275 | 0.608696    | 0.0     |
| 1541 | 0.333333   | 0.079019 | 0.077181 | 0.105175 | 0.913043    | 1.0     |
| 5384 | 0.166667   | 0.018529 | 0.026846 | 0.040341 | 0.521739    | 0.0     |

## E. Choosing and Training a model

- Linear Regression

```
from sklearn.linear_model import LinearRegression
```

```
reg_model = LinearRegression()
```

```
reg_model.fit(training_data_preprocessed, training_labels)
```

```
LinearRegression()
```

```
reg_model.coef_
```

```
array([-6.79884531e-02, -3.64027394e+00, -7.46086312e+01,  9.41891794e+01,
       -3.34135831e-01,  9.69987391e-01])
```

```
reg_model.intercept_
```

```
-1.4774092461326536
```

- SVM Regression

```
from sklearn.svm import LinearSVR, SVR

lin_svr = LinearSVR()
lin_svr.fit(training_data_preprocessed, training_labels)
```

```
LinearSVR()
```

```
from sklearn.metrics import mean_squared_error

predictions = reg_model.predict(training_data_preprocessed)
```

```
from sklearn.metrics import mean_squared_error

predictions = lin_svr.predict(training_data_preprocessed)
mse = mean_squared_error(training_labels, predictions)
rmse = np.sqrt(mse)
rmse
```

```
1.4654912062849366
```

```
mse
```

```
2.1476644756984786
```

```
mae = mean_absolute_error(training_labels, predictions)
```

```
mae
```

```
0.8405229840999318
```

- Random Forest

```
X_train_scaled = training_data_preprocessed
y_train = training_labels
X_train = train_data
```

```
from sklearn.ensemble import RandomForestRegressor

forest_reg = RandomForestRegressor(min_samples_split=2,bootstrap=False, random_state=42,n_jobs=-1)

forest_reg.fit(X_train_scaled, y_train)
```

```
RandomForestRegressor(bootstrap=False, n_jobs=-1, random_state=42)
```

```
from sklearn.metrics import mean_squared_error

def predict(input_data,model,labels):
    """
    Take the input data, model and labels and return predictions
    """

    preds = model.predict(input_data)
    mse = mean_squared_error(labels,preds)
    rmse = np.sqrt(mse)
    rmse

    return rmse
```

```
predict(X_train_scaled, forest_reg, y_train)
```

```
4.26597753845401e-15
```

- Decision Tree

```
from sklearn.tree import DecisionTreeRegressor

tree_reg_scaled = DecisionTreeRegressor()

tree_reg_scaled.fit(X_train_scaled, y_train)
```

```
DecisionTreeRegressor()
```



```

from sklearn.metrics import mean_squared_error

def predict(input_data,model,labels):
    """
    Take the input data, model and labels and return predictions
    """

    preds = model.predict(input_data)
    mse = mean_squared_error(labels,preds)
    rmse = np.sqrt(mse)
    rmse

    return rmse

```

```
predict(X_train_scaled, tree_reg_scaled, y_train)
```

```
4.09631747367848e-16
```

## F. Model Evaluation

### Comparing MAE, MSE, and RMSE

```

from sklearn.metrics import mean_absolute_error

mae = mean_absolute_error(training_labels, predictions)
mae

```

```
0.6179293317885355
```

```
mse
```

```
0.9441483520628701
```

```
rmse
```

```
0.9716729655922666
```

Diantara ketiga evaluation metric untuk train data, nilai **MAE** yang paling mendekati 0 yaitu `0.617` .

### Comparing prediction and the actual data

```
preds = pd.DataFrame(predictions, index = training_labels.index, columns=['pred'])
```

```
preds['act'] = training_labels # act means actual
```

```
preds.describe()
```

|       | pred        | act         |
|-------|-------------|-------------|
| count | 5789.000000 | 5789.000000 |
| mean  | 1.980869    | 1.980869    |
| std   | 2.271022    | 2.470192    |
| min   | -2.005195   | 0.000000    |
| 25%   | 0.282569    | 0.000000    |
| 50%   | 2.053546    | 1.660000    |
| 75%   | 2.830464    | 2.760000    |
| max   | 28.997355   | 33.200000   |

## G. Evaluating the Model on the Test Set

| test_preprocessed |            |          |          |          |             |         |
|-------------------|------------|----------|----------|----------|-------------|---------|
|                   | passengers | distance | fare     | total    | pickup_hour | payment |
| 2509              | 0.166667   | 0.028883 | 0.057047 | 0.072038 | 0.608696    | 1.0     |
| 5462              | 0.333333   | 0.104360 | 0.093960 | 0.105233 | 1.000000    | 1.0     |
| 4869              | 0.166667   | 0.146049 | 0.117450 | 0.146727 | 0.913043    | 1.0     |
| 5124              | 0.166667   | 0.049046 | 0.057047 | 0.072038 | 0.304348    | 1.0     |
| 756               | 0.833333   | 0.040327 | 0.060403 | 0.074919 | 0.782609    | 0.0     |
| ...               | ...        | ...      | ...      | ...      | ...         | ...     |
| 1980              | 0.166667   | 0.096185 | 0.093960 | 0.122522 | 0.869565    | 1.0     |
| 4209              | 0.166667   | 0.060763 | 0.083893 | 0.089327 | 0.434783    | 0.0     |
| 3258              | 0.166667   | 0.029973 | 0.036913 | 0.063105 | 0.434783    | 1.0     |
| 4906              | 0.166667   | 0.193460 | 0.164430 | 0.183379 | 0.304348    | 1.0     |
| 962               | 0.166667   | 0.250409 | 0.197987 | 0.248790 | 0.695652    | 1.0     |

644 rows × 6 columns

```
test_pred = reg_model.predict(test_preprocessed)
test_mse = mean_squared_error(test_labels, test_pred)

test_rmse = np.sqrt(test_mse)
test_rmse
```

0.965552259821685

test\_mse

0.9322911664467627

```
test_mae = mean_absolute_error(test_labels, test_pred)
```

test\_mae

0.5853923660362264

Diantara ketiga evaluation metric untuk test data, nilai **MAE** yang paling mendekati 0 yaitu **\$0.585**.

## H. Conclusion

```
con = pd.DataFrame({'mae': [mae, test_mae],
                    'mse': [mse, test_mse],
                    'rmse': [rmse, test_rmse]},
                    index=['train_data', 'test_data'])

con
```

|            | mae      | mse      | rmse     |
|------------|----------|----------|----------|
| train_data | 0.617929 | 0.944148 | 0.971673 |
| test_data  | 0.585392 | 0.932291 | 0.965552 |

Setelah melalui serangkaian langkah dalam pembangunan machine learning menggunakan model **Linear Regression**, didapati bahwa model yang dibangun sudah cocok dengan metode *normalization* untuk *feature scaling*, *ordinary* untuk *encoding*, dan evaluasi training data terbaik menggunakan **MAE** dengan hasil **\$0.617** karena lebih mendekati 0. Untuk test data evaluasi terbaik menggunakan **MAE** dengan hasil **\$0.585**. Hal ini berarti bahwa model yang kami gunakan merupakan best model karena testing error < training error.