

**LAPORAN AKHIR**  
**MAGANG & STUDI INDEPENDEN BERSERTIFIKAT**

**Data Science Basic**  
**Studi Independen**  
**Di Celerates (PT Mitra Talenta Grup)**

Ricky Ahmad Faizal  
623220020

Nama Dosen Pendamping Program (DPP):  
Dominikus Arif Budi Prasetyo, M.Si



**PROGRAM STUDI DIII MANAJEMEN INFORMATIKA**  
**UNIVERSITAS LOGISTIK DAN BISNIS INTERNASIONAL**  
**BANDUNG**  
**2024**



## Kata Pengantar

Puji syukur saya panjatkan kehadiran Tuhan Yang Maha Esa yang telah melimpahkan rahmat dan hidayah-Nya, sehingga saya dapat menyelesaikan program Studi Independen Bersertifikat Kampus Merdeka - Celerates - Data Science Basic dengan baik. Perjalanan dalam mengikuti program ini hingga selesai tidak terlepas dari berbagai pihak yang memberi kontribusi dari segi moril maupun materiil. Oleh sebab itu, saya ucapkan terima kasih kepada:

1. Kampus Merdeka yang memberikan kesempatan untuk belajar dengan para praktisi melalui programnya.
2. Celerates beserta jajarannya yang telah sukses menyelenggarakan program Celerates sebagai wadah aktualisasi diri para mahasiswa yang tertarik mengenai dunia digital.
3. Universitas Logistik dan Bisnis Internasional yang telah mengizinkan saya untuk mengikuti program pengembangan skills ini.
4. Trainer, Mentor, dan DPP yang telah kebersamai saya selama berproses dalam meningkatkan pengetahuan, keterampilan, dan nilai dalam diri melalui program ini.
5. Tim Celerates - Data Science Basic dan teman teman satu angkatan yang terus mendukung selama proses pembelajaran berlangsung.

Dengan selesainya Studi Independen Bersertifikat ini, saya berharap dapat mengembangkan kemampuan saya sehingga benar-benar menjadi “**Talenta Digital - Data Science Basic**” yang berkualitas di masa mendatang.

## Daftar Isi

Kata Pengantar .....	i
Daftar Isi .....	ii
Daftar Gambar .....	iii
I. Gambaran Umum.....	1
A. Profil Perusahaan.....	1
B. Deskripsi Kegiatan .....	1
1. Posisi .....	1
2. Deskripsi.....	1
3. Tujuan.....	2
4. Lingkup Pembelajaran & Proyek .....	3
5. Alur Program.....	4
6. Final Project .....	12
II. Aktivitas Bulanan .....	30
A. Sertifikasi Keahlian .....	39
III. Penutup.....	40
A. Kesimpulan.....	40
B. Saran .....	41
Referensi .....	
Lampiran .....	

## Daftar Gambar

Gambar 1. Load Data .....	17
Gambar 2. Analisis Dataset .....	17
Gambar 3. Kolom value null .....	18
Gambar 4. Analisis deskripsi tabel.....	18
Gambar 5. Model KNN .....	20
Gambar 6. Model Decision Tree Classifier.....	21
Gambar 7. Model Support Vector Model.....	22
Gambar 8. Model Gradient Boosting Classifier.....	22
Gambar 9. Sintaks Visualisasi Data .....	23
Gambar 10. Sintaks Visualisasi Data .....	24
Gambar 11. Visualisasi Data .....	25



## **I. Gambaran Umum**

### **A. Profil Perusahaan**

Celerates adalah perusahaan yang startup dengan basis IT Consultant. Celerates dapat memberikan solusi bagi permasalahan IT dan pengembangan teknologi bagi perusahaan yang ingin mengembangkan digitalisasi industri.

Celerates menyediakan tenaga-tenaga IT, yang memiliki kemampuan dan kompetensi yang memadai. Talent IT di Celerates disiapkan dengan berbagai teknologi modern yang digunakan dalam industri 4.0.

Celerates didirikan untuk membantu perusahaan-perusahaan di Indonesia mengembangkan teknologi industri, menyesuaikan dengan berkembangnya digitalisasi di Indonesia.

Kami sadar, kebutuhan para pekerja IT saat ini cukup besar, tapi tidak semua perusahaan faham apa yang mereka butuhkan untuk membuat sebuah teknologi baru. Kami hadir sebagai solusi digitalisasi bagi perusahaan Anda, dengan memberikan ke arah mana perusahaan Anda ingin mengembangkan teknologinya, atau apa yang ingin Anda lakukan ke depannya.

Satu hal yang ingin kami utamakan, kami tau tenaga IT di Indonesia memiliki kemampuan yang tidak kalah dengan tenaga IT dari luar negeri, dan Indonesia memiliki perkembangan IT yang tidak kalah dengan negara-negara lain. Itu yang membuat kami bersemangat untuk mengembangkan teknologi dan juga IT di Indonesia.

### **B. Deskripsi Kegiatan**

#### **1. Posisi**

Peserta Celerates Batch 6 - Data Science Basic

#### **2. Deskripsi**

Celerates Batch 6 - Data Science Basic merupakan Studi Independen Bersertifikat yang diselenggarakan oleh PT Mitra Talenta Grup dan didukung penuh oleh Kemdikbud Ristekdikti di bawah naungan inisiatif KAMPUS MERDEKA. Batch 6 diselenggarakan dari Februari hingga Juni 2024. Program ini dirancang sepenuhnya secara daring untuk mempersiapkan talenta digital - Data Science Basic yang mampu bersaing di era digital.

Ruang lingkup Program meliputi:

- Pembekalan materi pelatihan secara **daring**;
- **Pendampingan** dengan mentor dan fasilitator;
- **Networking** dengan berbagai partner dalam dunia digital;
- **Collaborating** dalam menyelesaikan *project* bersama peserta lain untuk mendapatkan *real-life experience*.

Fasilitas yang akan diberikan kepada tim peserta meliputi:

- Kurikulum pembelajaran dirancang oleh pengajar terbaik dalam ekosistem digital;
- Konversi kredit 20 SKS;
- *Mentorship* intensif dengan mentor dari ekosistem digital;
- *Networking* dengan pemain penting dalam ekosistem digital;
- Sertifikat Kompetensi Digital (*Data Science Basic*)
- *Learning Management System*;
- Modul dan referensi belajar terbaik; dan
- Pengalaman berkolaborasi dalam menyelesaikan *project* bersama.

Tim terpilih akan mendapatkan kurikulum pembelajaran selama program, dengan capaian pembelajaran yang akan dikonversikan ke dalam SKS kurikulum masing-masing kampus asal mahasiswa. Untuk bisa menjalankan program dengan baik, tim peserta akan mendapatkan pembekalan materi dan pendampingan intensif oleh mentor profesional selama program berlangsung.

### 3. Tujuan

Peserta diharapkan memiliki kompetensi sebagai talenta digital (*Data Science Basic*), yang mampu berkolaborasi dan mengerjakan real project sehingga siap berkontribusi dan berkarier di industri digital. Secara lebih rinci, setelah menyelesaikan studi independen, peserta diharapkan mampu:

- mengenal *data science* secara umum;
- mengaplikasikan bahasa pemrograman SQL dalam mengolah data;
- mengaplikasikan bahasa pemrograman *Python* sebagai dasar dalam *data science*;



- memahami konsep statistika sebagai dasar dalam *data science*;
- memahami cara mempersiapkan data;
- memproses data mentah menjadi data yang siap digunakan;
- memahami konsep *machine learning* (*supervised* dan *unsupervised*);
- melakukan berbagai model *machine learning* dalam menyelesaikan suatu masalah;
- mengevaluasi model dari *machine learning* dengan metode yang sesuai;
- memahami cara memvisualisasikan dan menyampaikan hasil olah data;
- membuat dashboard interaktif hasil analisis data;

#### **4. Lingkup Pembelajaran & Proyek**

Selama menjalani program ini, peserta menyelesaikan beberapa tugas berdasarkan materi-materi yang diberikan melalui *Learning Management System (LMS)*. Tugas ini terdiri dari tugas individu dan tugas tim. Lingkup pembelajaran dan proyek meliputi:

- *Introduction to Data Science; SQL Foundation; Basic Python Foundation; Advanced Python Foundation; Statistics for Data Science.*
- *Data Preparation: Transforming, Cleansing, Merging, etc; Exploratory Data Analysis (EDA); Feature Engineering.*
- *Introduction to Machine Learning; Unsupervised Learning - Clustering; Supervised Learning - Regression; Supervised Learning - Classification: Decision Trees, Random Forest, Support Vector Machine; Time Series Forecasting; Cross Validation & Learning Curves; Model Optimization: Hyperparameter Tuning & Bootstrapping; Metric Evaluation.*
- *Data Visualization: Working with Tableau Tools; Data Storytelling and Communication.*
- *Final Project and Showcase*

## 5. Alur Program

Alur program secara umum dijalankan melalui beberapa acuan di bawah ini.

### a. Kompetensi Inti

Terdapat beberapa kompetensi inti yang ingin dicapai dan dikembangkan mencakup:

1. **Communication:** *Assertive (capable of providing feedback and delivering own thoughts), public speaking.*
2. **Problem solving:** *Critical thinker, thinks things through, open-minded, strategic, tactical.*
3. **Social Intelligence:** *Approachability, helpfulness, courtesy, and empathy.*
4. **Creativity:** *Original & adaptive, clever, sess and does things in different ways.*

### b. Proses Induksi

1. Pengenalan program, alur program, hasil yang diharapkan, serta skema penilaian bagi peserta.
2. Pengenalan terhadap mentor profesional yang akan mendampingi para peserta selama program berjalan.

### c. Sistem Pembelajaran

Sistem pembelajaran di Program Data Science Basic terdiri atas:

1. **Pembelajaran asinkronus:** Peserta wajib mempersiapkan pembelajaran mandiri dengan memproses seluruh materi yang diberikan melalui video, bahan bacaan, tugas, serta referensi lainnya sebelum pembelajaran sinkronus. Materi dapat diakses melalui *Learning Management System (LMS)*.
2. **Pembelajaran sinkronus:** Peserta akan mengikuti pembelajaran aktif melalui kelas sinkronous dengan para *expert* di bidangnya berbentuk zoom meet, diskusi, dan lain sebagainya.

#### **d. Sistem Penilaian**

Proses penilaian mahasiswa dilaksanakan oleh seluruh pihak termasuk mentor, dan sesama peserta (*peer to peer evaluation*). Penilaian mahasiswa dilakukan melalui beberapa metode:

##### **1. Tugas Individu (Belajar Mandiri & Kuis)**

Peserta diberikan buku dan referensi belajar melalui e-book, artikel, case study, video, dan sebagainya. Sebelum mengikuti kelas sinkronus, peserta wajib membaca referensi yang sudah diberikan, lebih lanjut, terdapat kuis untuk menguji pemahaman peserta.

##### **2. Keaktifan selama Proses Pembelajaran**

Pembelajaran sinkronus adalah sesi interaktif berdurasi 2-3 jam dengan *expert speaker* yang merupakan ahli atau praktisi bidang *Data Science Basic*. Pada sesi ini, peserta akan melakukan diskusi, *problem solving*, simulasi, presentasi, serta aktivitas lainnya yang mendorong peserta aktif selama sesi tersebut. Selama proses *live session*, para *expert* akan melihat keaktifan dan pemahaman materi mahasiswa.

##### **3. Mentorship**

Selama program, tiap tim akan mendapatkan *mentorship* dari mentor penanggung jawab dan fasilitator. Mentor penanggung jawab akan berdedikasi penuh dalam pengembangan dari sisi akademis dan praktikal peserta dari awal hingga akhir program. Fasilitator adalah pembimbing yang bertanggung jawab untuk memastikan peserta mengikuti pembelajaran dengan maksimal. Peran fasilitator termasuk membantu peserta merancang jadwal belajar mingguan, mengingatkan tugas dan proyek yang harus dikerjakan, mengingatkan jadwal kelas, dan kegiatan lainnya.

#### **4. Proyek Tim (*Final Project*)**

Peserta akan mengerjakan berbagai proyek tim yang bertujuan untuk mengasah kemampuan *softskill* maupun *hardskill*. Peserta akan mengerjakan proyek akhir dengan kompetensi *Data Science Basic* dan berkolaborasi dengan peserta studi independen lainnya. Proyek tim ini diharapkan akan memberikan kondisi *real* di industri. Peserta dibebaskan untuk menginisiasi proyek yang ingin dikerjakan sesuai *dataset* dan batasan yang diberikan. Proyek harus memiliki nilai bisnis.

#### **5. *Showcase***

Masing-masing tim akan mempresentasikan proyek final tim yang telah disusun di depan para juri. Tim terbaik akan dipilih untuk mendapatkan sertifikat dan hadiah sebagai bentuk apresiasi.

#### **e. Pembelajaran Sesuai Kurikulum**

Peserta mendapatkan kurikulum pembelajaran selama program, dengan capaian pembelajaran yang akan dikonversikan ke dalam SKS kurikulum masing-masing kampus asal mahasiswa. Tiap peserta juga akan mendapatkan pendampingan intensif dari para mentor profesional selama program berlangsung.

Berikut adalah kurikulum Program Data Science Basic.

##### **1. *Foundation for Data Science***

###### ***Targeted Skill Level***

Mahasiswa mampu mengenal *data science* secara umum; mampu mengaplikasikan bahasa pemrograman SQL dalam mengolah data; mampu mengaplikasikan bahasa pemrograman *Python* sebagai dasar dalam *data science*; mampu memahami konsep statistika sebagai dasar dalam *data science*.

###### ***Learning Details***

Di chapter *Foundation for Data Science*, peserta akan belajar materi-materi dasar terkait pilar sebagai seorang *data scientist*.

Adapun detail materi pembelajaran mencakup subchapter sebagai berikut.

- *Introduction to Data Science*
- *SQL Foundation*
- *Basic Python Foundation*
- *Advanced Python Foundation*
- *Statistics for Data Science*

Chapter ini akan berisi teori dan hands on yang dilakukan dengan *Google Colaboratory*..

Chapter ini dilengkapi dengan pembelajaran asinkronus melalui modul, *learning video*, *quiz*, *exercise*, dan referensi pembelajaran lain di *Learning Management System* yaitu Notion Celerates

Pembelajaran sinkronus dilakukan melalui pemaparan materi, diskusi, dan tanya jawab pada sesi *zoom meet* dan mentoring. Pemahaman materi didukung juga oleh sesi 'belajar bareng' bersama '*study buddy*'.

Proses pembelajaran diperkaya dengan assignment untuk mengasah kemampuan analisis dan evaluasi mengenai materi dasardasar sebagai data scientist menggunakan Python dan SQL.

#### ***Learning Duration***

135 hours

#### ***Learning Resources***

*Modul* dan *exercise* dari Celerates, *learning video* dan kuis yang disusun oleh *learning designer* Celerates, *studium generale*, referensi lainnya yang dapat diakses melalui *Notion*.

#### ***Assessment Method***

Tugas individu dan/atau kelompok, keaktifan, kuis

## **2. *Data Preprocessing***

### ***Targeted Skill Level***

Mahasiswa mampu memahami cara mempersiapkan data; mampu memproses data mentah menjadi data yang siap digunakan.

### ***Learning Details***

Di chapter *Data Preprocessing*, peserta akan belajar terkait materi persiapan dalam mengolah data mentah hingga menjadi data yang kaya akan fitur-fitur yang bermanfaat. Adapun detail materi pembelajaran mencakup subchapter sebagai berikut.

- *Data Preparation: Transforming, Cleansing, Merging, etc*
- *Exploratory Data Analysis (EDA)*
- *Feature Engineering*

Chapter ini akan berisi teori dan hands on yang dilakukan dengan *Google Colaboratory*.

Chapter ini dilengkapi dengan pembelajaran asinkronus melalui modul, *learning video*, *quiz*, *exercise*, dan referensi pembelajaran lain di *Learning Management System* yaitu Notion Celerates

Pembelajaran sinkronus dilakukan melalui pemaparan materi, diskusi, dan tanya jawab pada zoom meet *mentoring*. Pemahaman materi didukung juga oleh sesi 'belajar bareng' bersama '*study buddy*'.

Proses pembelajaran diperkaya dengan assignment untuk mengasah kemampuan analisis dan evaluasi mengenai pengolahan data.

### ***Learning Duration***

135 hours

### ***Learning Resources***

*Modul* dan *exercise* dari Celerates *learning video* dan kuis yang disusun oleh *learning designer* Celerates, *studium generale*,

dan referensi lainnya yang dapat diakses melalui *Learning Management System*.

#### ***Assessment Method***

Tugas individu dan/atau kelompok, keaktifan, kuis

### **3. *Machine Learning***

#### ***Targeted Skill Level***

Mahasiswa mampu memahami konsep *machine learning* dan *Deep Learning* (*supervised* dan *unsupervised*) serta mampu melakukan berbagai model *machine learning* dalam menyelesaikan suatu masalah. Diharapkan pula mampu mengevaluasi model dari *machine learning* dengan metode yang sesuai.

#### ***Learning Details***

Di chapter *Machine Learning*, peserta akan belajar berbagai macam model dan algoritma *machine learning* untuk mengolah data dan sebagai dasar untuk menjadi data science. Peserta juga akan mempelajari cara mengevaluasi model *machine learning* yang sudah dibuat. Adapun detail materi pembelajaran mencakup subchapter sebagai berikut.

- *Introduction to Machine Learning*
- *Unsupervised Learning - Clustering*
- *Supervised Learning - Regression*
- *Supervised Learning - Classification: Decision Trees, Random Forest, Support Vector Machine*
- *Time Series Forecasting*
- *Cross Validation & Learning Curves*
- *Model Optimization: Hyperparameter Tuning & Bootstrapping*
- *Metric Evaluation*

Chapter ini akan berisi teori dan hands on yang dilakukan dengan *Google Colaboratory*.

Chapter ini dilengkapi dengan pembelajaran asinkronus melalui modul, *learning video*, *quiz*, *exercise*, dan referensi pembelajaran lain di *Learning Management System* yaitu Notion.

Pembelajaran sinkronus dilakukan melalui pemaparan materi, diskusi, dan tanya jawab pada sesi *studium generale*, *mandatory mentoring*, dan *ad hoc mentoring*. Pemahaman materi didukung juga oleh sesi 'belajar bareng' bersama '*study buddy*'.

Proses pembelajaran diperkaya dengan assignment untuk mengasah kemampuan terkait analisis, evaluasi, dan kemampuan menciptakan serta memodifikasi berbagai model machine learning untuk menyelesaikan permasalahan yang ada.

#### ***Learning Duration***

180 hours

#### ***Learning Resources***

*Modul* dan *exercise* dari Celerates, *learning video* dan kuis yang disusun oleh *learning designer* Celerates, *studium generale*, dan referensi lainnya yang dapat diakses melalui *Learning Management System*.

#### ***Assessment Method***

Tugas individu dan/atau kelompok, keaktifan, kuis

### **4. Data Visualization**

#### ***Targeted Skill Level***

Mahasiswa mampu memahami cara memvisualisasikan dan menyampaikan hasil olah data; mampu membuat *dashboard* interaktif hasil analisis data.

#### ***Learning Details***

Di chapter *Data Visualization*, peserta akan belajar materi visualisasi data menggunakan *BI Tools*, seperti *Tableau* beserta cara menyampaikan hasil visualisasi yang baik. Adapun detail materi pembelajaran mencakup subchapter sebagai berikut.

- *Data Visualization: Working with BI Tools based on*



### *Gestalt Principle*

- *Data Storytelling and Communication*

Chapter ini akan berisi teori dan hands on yang dilakukan dengan *Google Colaboratory*, *Looker Studio*, dan *Tableau*.

Chapter ini dilengkapi dengan pembelajaran asinkronus melalui modul, *learning video*, *quiz*, *exercise*, dan referensi pembelajaran lain di *Learning Management System* yaitu Notion Celerates

Pembelajaran sinkronus dilakukan melalui pemaparan materi, diskusi, dan tanya jawab pada sesi *studium generale*, *mandatory mentoring*, dan *adhoc mentoring*. Pemahaman materi didukung juga oleh sesi 'belajar bareng' bersama 'study buddy'.

Proses pembelajaran diperkaya dengan assignment untuk mengasah kemampuan analisis, evaluasi, dan kemampuan membuat visualisasi data serta mengkomunikasikan insight dari data.

#### ***Learning Duration***

90 hours

#### ***Learning Resources***

*Modul* dan *exercise* dari Celerates, *learning video* dan kuis yang disusun oleh *learning designer* Celerates, *studium generale*, dan referensi lainnya yang dapat diakses melalui *Learning Management System*.

#### ***Assessment Method***

Tugas individu dan/atau kelompok, keaktifan, kuis

## **5. Final Project & Showcase**

#### ***Targeted Skill Level***

Mahasiswa mampu menyelesaikan permasalahan bisnis menggunakan pemodelan kuantitatif dan teknik analisis data mampu mengkomunikasikan hasil analisis atau mempresentasikan hasil visualisasi data dengan efektif.

### ***Learning Details***

Pada tahap ini, peserta akan berkolaborasi bersama tim untuk membuat produk *data science*. Peserta akan diberi kebebasan untuk memilih dan berkreasi dalam membuat produk yang diminati dengan pendekatan-pendekatan yang telah diajarkan di kelas dengan berbagai sudut pandang dalam berbagai studi kasus. Peserta akan melakukan analisis dan menyusun hasil temuan berdasarkan kasus. Kemudian, merancang dan menjalankan eksperimen pada berbagai hipotesis untuk fitur baru, merumuskan dan menyajikan kesimpulan akhir, atau membuat model berdasarkan data yang ada.

Selama proses pengerjaan final project, peserta akan didampingi oleh expert dan mentor. Pengerjaan final project ini akan dilombakan. Di akhir program, setiap tim akan mempresentasikan produknya di depan para expert dan peserta lainnya.

### ***Learning Duration***

270 hours

### ***Learning Resources***

*Modul dan exercise* dari Celerates, *learning video* dan kuis yang disusun oleh *learning designer* Celerates, *studium generale*, dan referensi lainnya yang dapat diakses melalui *Learning Management System*.

### ***Assessment Method***

Tugas individu dan/atau kelompok, keaktifan, kuis

## **6. Final Project**

### **a. Latar Belakang Final Project**

Perkembangan *data science basic* saat ini mengalami perkembangan yang pesat. Dalam berbagai sektor di kehidupan, setiap harinya, milyaran data terproduksi. Dari data-data tersebut, dibutuhkan penerapan *data science* untuk mendapatkan *insight* dan solusi untuk memecahkan masalah untuk mempermudah pekerjaan dengan teknologi masa kini. Sebagai calon talenta digital, peserta Program Data Science diharapkan mampu

menyelesaikan permasalahan dalam *final project* layaknya seorang *data scientist*.

*Final Project* Program Data Science Basic Celerates Batch 6 adalah proyek akhir yang dirancang oleh *subject matter expert* Program Data Science yaitu para trainer dan para mentor. Peserta diminta untuk membuat proyek akhir dengan nilai bisnis sesuai dengan kesepakatan masing-masing, mulai dari pemilihan *study case* hingga narasi dengan *dataset* yang diberikan. Pengerjaan *final project* berlangsung selama kurang lebih 2 bulan dan dikerjakan dalam tim beranggotakan 4-5 orang.

Topik *final project* yang dapat dipilih oleh peserta Program Data Science Basic adalah sebagai berikut.

### 1. *Classification*

Teknik ini digunakan untuk mengelompokkan data ke dalam kategori atau kelas yang berbeda. Ini sering digunakan untuk memahami perilaku sumber data yang diamati. *Classification* terdiri dari *binary classification* dan *multivariate classification*.

### 2. *Clustering*

Teknik ini digunakan untuk mengelompokkan data ke dalam kelompok atau klaster berdasarkan kesamaan mereka. Ini membantu dalam mengidentifikasi pola yang mungkin tidak terlihat secara langsung dan dapat digunakan untuk membuat segmen data berdasarkan perilaku atau karakteristik tertentu.

### 3. *Regression*

Teknik ini digunakan untuk memprediksi nilai numerik kontinu (misalnya, harga rumah, gaji karyawan) berdasarkan satu atau lebih variabel independen (misalnya, luas rumah, pengalaman kerja). Teknik ini menggunakan model statistik untuk mempelajari hubungan antara variabel independen dan dependen, dan kemudian

menggunakan model tersebut untuk memprediksi nilai dependen untuk data baru.

#### 4. *Time Series Forecasting*

Teknik ini digunakan untuk memprediksi nilai masa depan dari suatu deret waktu (time series) berdasarkan data historis. Teknik ini menggunakan model statistik atau algoritma machine learning untuk mempelajari pola dan tren dalam deret waktu, dan kemudian menggunakan model tersebut untuk memprediksi nilai di masa depan.

#### 5. *Recommendation System*

Teknik ini digunakan untuk merekomendasikan item (misalnya, produk, film, musik) kepada pengguna berdasarkan preferensi mereka. Teknik ini menggunakan model machine learning untuk mempelajari preferensi pengguna dari data historis, seperti interaksi pengguna dengan item (misalnya, pembelian, rating, klik), dan kemudian menggunakan model tersebut untuk merekomendasikan item yang kemungkinan besar disukai pengguna.

Peserta juga dibebaskan menggunakan teknik lain di luar teknik yang telah disebutkan di atas dengan berdiskusi dengan mentor.

Setelah waktu pengerjaan *final project* selesai, hasil *final project* akan dilombakan dalam *Selection to Showcase* untuk mencari 10 tim terbaik yang akan maju ke babak *Showcase* dan berkesempatan memenangkan hadiah dari

#### **b. Proses Pengerjaan Final Project**

Berikut adalah uraian proses pengerjaan *final project*.

## 1. Pembentukan Tim

Pembentukan tim *final project* dibagi berdasarkan latar belakang universitas dan program studi, keaktifan, level teknikal dan *softskill* peserta.

Kelompok 3 Mentor Asfaar terdiri atas:

- a. M. Fadhilah Syahri Kamil
- b. Amalia Husna Sentosa
- c. Farah Meishinthya
- d. Ricky Ahmad Faizal
- e. Ramadhani Renanthera Taritia

Sebelum memulai pengerjaan *final project*, telah dilaksanakan proses diskusi di Gmeet menyepakati tugas tugas yang diberikan

## 2. Penentuan Dataset, Topik, & Judul *Final Project*

Setelah pembentukan tim, dilaksanakan penentuan dataset, topik, dan judul *final project*. Topik dan judul harus didasarkan pada permasalahan nyata yang ingin diselesaikan dan memiliki nilai bisnis. Dataset, topik, dan judul harus diajukan kepada mentor untuk disetujui.

Berikut pilihan topik dan dataset yang dapat dipilih:

- *Topik:* Employee Churn Rate Prediction “Leave or Not”
- *Dataset:* Employee Dataset

Dataset yang dipilih dalam *final project* adalah *Dataset 1: Employee Dataset*. Dalam *final project* ini dipilih topik forecasting dan clustering dengan judul “Analisis Regional Super App Memahami Pola Konsumsi Dan Preferensi Produk Di Beberapa Provinsi Indonesia”.

Latar belakang permasalahan diambilnya judul tersebut adalah bagaimana perusahaan Super App dapat memahami dan mengidentifikasi tren serta pola penggunaan aplikasinya yang bervariasi di setiap provinsi di Indonesia, sehingga dapat menyesuaikan operasional dan strategi bisnis untuk memenuhi kebutuhan yang beragam dari pengguna di setiap provinsi dengan lebih efektif dan efisien.

Proyek/produk yang dianalisis memiliki target pengguna yaitu Perusahaan *Super App* dan bertujuan untuk menerapkan analisis data yang mendalam untuk memahami dinamika regional, personalisasi pengalaman pengguna, dan mengidentifikasi peluang pengembangan produk. Dengan pendekatan berbasis data ini, perusahaan Super App dapat meningkatkan efektivitas operasional, menyusun strategi yang lebih tepat, dan terus berinovasi untuk memenuhi kebutuhan pasar yang dinamis dan beragam di seluruh Indonesia.

Dataset yang diambil oleh tim dalam *project* ini adalah *dataset 1*.

Terdapat *file csv* didalam *dataset* ini dijelaskan sebagai berikut beserta dengan penjelasan setiap *file*:

- a. *employee.csv*: Berisi informasi terperinci mengenai Karyawan, meliputi data dan riwayat karyawan yang membantu memahami perilaku pengguna. Meliputi variabel: Education', 'JoiningYear', 'City', 'PaymentTier', 'Age', 'Gender', 'EverBenched', 'ExperienceInCurrentDomain', 'LeaveOrNot'.

### 3. *Data Preprocessing*

#### 3.1 *Load Data*

Pada tahap awal, yaitu dilakukannya dengan menghubungkan google collab dengan google drive untuk mengambil dataset yang telah di upload di google drive. Library yang digunakan pertama kali yaitu pandas, fungsi pandas di gambar ini yaitu untuk membaca file csv kedalam dataframe.

```
▼ Pengolahan Data

from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

import pandas as pd

df = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/final/Employee.csv')
```

*Gambar 1. Load Data*

#### 3.1.2 *Analisis Dataset*

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4653 entries, 0 to 4652
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Education              4653 non-null  object
1   JoiningYear            4653 non-null  int64
2   City                  4653 non-null  object
3   PaymentTier           4653 non-null  int64
4   Age                   4653 non-null  int64
5   Gender                4653 non-null  object
6   EverBenched           4653 non-null  object
7   ExperienceInCurrentDomain 4653 non-null  int64
8   LeaveOrNot            4653 non-null  int64
dtypes: int64(5), object(4)
memory usage: 327.3+ KB

df.columns

Index(['Education', 'JoiningYear', 'City', 'PaymentTier', 'Age', 'Gender',
      'EverBenched', 'ExperienceInCurrentDomain', 'LeaveOrNot'],
      dtype='object')

df.shape

(4653, 9)
```

*Gambar 2. Analisis Dataset*

Pada tahap ini, dilakukannya analisis dataset yang bermula dari melihat nama nama kolom, jumlah kolom, dan data type nya. Pada dataset ini diketahui bahwa ada 9 kolom yaitu 'Education', 'JoiningYear', 'City', 'PaymentTier', 'Age', 'Gender', 'EverBenched', 'ExperienceInCurrentDomain', 'LeaveOrNot' dan pada jumlah baris yaitu berjumlah 4653.

### 3.1.3 Cek kolom value Null

```
[ ] df.isnull().sum()
Education      0
JoiningYear    0
City           0
PaymentTier    0
Age            0
Gender         0
EverBenched    0
ExperienceInCurrentDomain  0
LeaveOrNot      0
dtype: int64
```

Gambar 3. Kolom value null

Pada tahap ini dilakukannya pengecekan kolom yang value nya null. Dan alhasil tidak terdapat kolom yang menunjukan kolom nya null. Bisa dilihat dari gambar bahwa semua kolom memiliki hasil angka nol yang berarti kolom kolom tersebut tidak memiliki value null. Jika salah satu kolom menunjukan angka 1 saja berarti kolom tersebut mempunyai value yang null.

### 3.1.4 Analisis Deskripsi Tabel

```
df.describe()
```

	JoiningYear	PaymentTier	Age	ExperienceInCurrentDomain	LeaveOrNot
count	4653.000000	4653.000000	4653.000000	4653.000000	4653.000000
mean	2015.062970	2.698259	29.393295	2.905652	0.343864
std	1.863377	0.561435	4.826087	1.558240	0.475047
min	2012.000000	1.000000	22.000000	0.000000	0.000000
25%	2013.000000	3.000000	26.000000	2.000000	0.000000
50%	2015.000000	3.000000	28.000000	3.000000	0.000000
75%	2017.000000	3.000000	32.000000	4.000000	1.000000
max	2018.000000	3.000000	41.000000	7.000000	1.000000

Gambar 4. Analisis deskripsi tabel

**JoiningYear:** Tahun bergabungnya karyawan dengan perusahaan

**PaymentTier:** Tingkat pembayaran gaji karyawan

**Age:** Usia karyawan



**ExperienceInCurrentDomain:** Pengalaman karyawan di bidang saat ini

**LeaveOrNot:** Apakah karyawan tersebut keluar dari perusahaan atau tidak  
(0: tidak keluar, 1: keluar)

### ***3.1.5 Analisis Statistik***

#### **1. Joining Year**

Dari statistik JoiningYear, dapat disimpulkan bahwa rata-rata karyawan bergabung dengan perusahaan pada tahun 2015. Sebagian besar karyawan (75%) bergabung dengan perusahaan setelah tahun 2014. Dan karyawan yang paling awal bergabung yaitu di tahun 2012 dan paling akhir bergabung sesuai data yaitu di tahun 2018.

#### **2. Payment Tier**

Dari statistik PaymentTier, dapat disimpulkan bahwa rata-rata karyawan berada di tingkat pembayaran gaji 2. Sebagian besar karyawan (75%) berada di tingkat pembayaran gaji 3. Dan karyawan yang pembayaran gaji nya paling minim yaitu di angka 1 dan pembayaran gaji paling tinggi di angka 3.

#### **3. Age**

Dari statistik Age, dapat disimpulkan bahwa rata-rata usia karyawan adalah 29 tahun. Sebagian besar karyawan (75%) berusia di bawah 32 tahun. Dan karyawan yang paling muda yaitu di umur 22 dan paling tua yaitu di umur 41.

#### **4. Experience In Current Domain**

Dari statistik ExperienceInCurrentDomain, dapat disimpulkan bahwa rata-rata karyawan memiliki pengalaman 2.9 tahun di bidang saat ini. Sebagian besar karyawan (75%) memiliki pengalaman di bidang saat ini selama 4 tahun atau kurang.

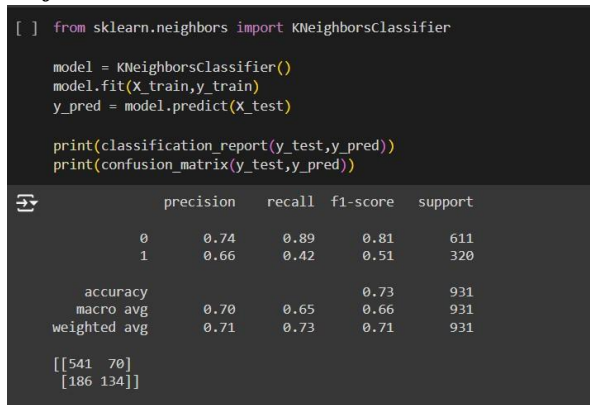
#### **5. Leave Or Not**

Dari statistik LeaveOrNot, dapat disimpulkan bahwa sekitar 34% karyawan keluar dari perusahaan.

## 4. Pemodelan dan Evaluasi

Tujuan dari pemodelan data dalam machine learning adalah untuk membuat representasi matematis dari data yang dapat digunakan untuk membuat prediksi atau pengambilan keputusan otomatis. Pemodelan data bertujuan untuk menangkap pola, hubungan, dan struktur dalam dataset sehingga model dapat menggeneralisasi informasi tersebut untuk diterapkan pada data baru atau tidak terlihat sebelumnya. Dengan kata lain, pemodelan data memungkinkan sistem machine learning untuk belajar dari data historis, memahami tren atau karakteristik penting, dan kemudian menerapkan pengetahuan tersebut untuk memberikan prediksi yang akurat, klasifikasi, atau rekomendasi di dunia nyata. Pemodelan yang efektif dapat meningkatkan efisiensi, akurasi, dan keandalan aplikasi machine learning dalam berbagai domain, mulai dari kesehatan dan keuangan hingga teknologi dan ilmu sosial. Di modelling ini menggunakan 4 model yaitu *KNeighborClassifier*, *DecisionTreeClassifier*, *Support Vector Machine* (SVM), dan terakhir yaitu *Gradient Boosting Classifier*.

### 1. *KNeighborClassifier*



Gambar 5. Model KNN

Pada algoritma KNN dapat diklasifikasikan data dengan cukup baik. Accuracy model adalah 0,73, yang berarti bahwa 70% data diklasifikasikan dengan benar. Precision untuk kelas 0 lebih tinggi daripada precision untuk kelas 1, yang berarti bahwa model

lebih baik dalam mengklasifikasikan data sebagai kelas 0. Recall untuk kelas 0 juga lebih tinggi daripada recall untuk kelas 1, yang berarti bahwa model lebih baik dalam mengidentifikasi data yang sebenarnya termasuk dalam kelas 0.

## 2. *DecisionTreeClassifier*

```
from sklearn.tree import DecisionTreeClassifier

model = DecisionTreeClassifier()
model.fit(X_train,y_train)
y_pred = model.predict(X_test)

print(classification_report(y_test,y_pred))
print(confusion_matrix(y_test,y_pred))
```

		precision	recall	f1-score	support
	0	0.77	0.91	0.83	611
	1	0.74	0.47	0.58	320
	accuracy			0.76	931
	macro avg	0.75	0.69	0.71	931
	weighted avg	0.76	0.76	0.75	931

```
[[558  53]
 [169 151]]
```

*Gambar 6. Model Decision Tree Classifier*

Pada Model Decision Tree Classifier dapat mengklasifikasikan data dengan cukup baik. Accuracy model adalah 0.75, yang berarti bahwa 75% data diklasifikasikan dengan benar. Precision dan recall untuk kedua kelas cukup tinggi, menunjukkan bahwa model dapat mengklasifikasikan data dengan baik dan mengidentifikasi data yang sebenarnya termasuk dalam kelas tersebut.

## 3. *Support Vector Machine (SVM)*

```

from sklearn.svm import SVC
from sklearn.metrics import accuracy_score

model = SVC(kernel='linear')
model.fit(X_train,y_train)
y_pred = model.predict(X_test)
model_acc = accuracy_score(y_pred, y_test)

print(classification_report(y_test,y_pred))
print(confusion_matrix(y_test,y_pred))
print("Akurasi SVM : {:.2f} %".format(model_acc*100))

```

	precision	recall	f1-score	support
0	0.68	0.90	0.77	611
1	0.51	0.21	0.29	320
accuracy			0.66	931
macro avg	0.60	0.55	0.53	931
weighted avg	0.62	0.66	0.61	931

```

[[547  64]
 [254  66]]
Akurasi SVM : 65.84 %

```

*Gambar 7. Model Support Vector Model*

Pada algoritma SVM dapat mengklasifikasikan data dengan cukup baik. Accuracy model adalah 0.60, yang berarti bahwa 60% data diklasifikasikan dengan benar. Precision untuk kelas 0 lebih tinggi daripada precision untuk kelas 1, yang berarti bahwa model lebih baik dalam mengklasifikasikan data sebagai kelas 0. Recall untuk kelas 0 juga lebih tinggi daripada recall untuk kelas 1, yang berarti bahwa model lebih baik dalam mengidentifikasi data yang sebenarnya termasuk dalam kelas 0.

#### 4. Gradient Boosting Classifier

```

from sklearn.ensemble import GradientBoostingClassifier

model = GradientBoostingClassifier()
model.fit(X_train,y_train)
y_pred = model.predict(X_test)

print(classification_report(y_test,y_pred))
print(confusion_matrix(y_test,y_pred))

```

	precision	recall	f1-score	support
0	0.78	0.95	0.85	611
1	0.84	0.47	0.61	320
accuracy			0.79	931
macro avg	0.81	0.71	0.73	931
weighted avg	0.80	0.79	0.77	931

```

[[581  30]
 [168 152]]

```

*Gambar 8. Model Gradient Boosting Classifier*

Berdasarkan hasil klasifikasi, diperoleh nilai akurasi sebesar 80%. Nilai akurasi ini menunjukkan bahwa algoritma Gradient Boosting Classifier mampu mengklasifikasikan data dengan baik. Nilai precision dan recall untuk kelas 0 masing-masing adalah 0.78

dan 0.95. Nilai ini menunjukkan bahwa algoritma Gradient Boosting Classifier mampu mengidentifikasi kelas 0 dengan baik. Nilai precision dan recall untuk kelas 1 masing-masing adalah 0.84 dan 0.47. Nilai ini menunjukkan bahwa algoritma Gradient Boosting Classifier kurang baik dalam mengidentifikasi kelas 1.

Dapat disimpulkan bahwa model *K-Neighbor Classifier* yaitu dengan total nilai *accuracy*-nya yaitu di 0.73. Lalu yang kedua menggunakan model *Decision Tree Classifier* yaitu dengan total nilai *accuracy*-nya yaitu di 0.76. Selanjutnya yang ketiga dengan model *Support Vector Machine* (SVM) yaitu dengan total nilai *accuracy*-nya yaitu di 0.66. Dan yang terakhir menggunakan model *Gradient Boosting Classifier* yaitu dengan total nilai *accuracy*-nya yaitu 0.79.

## 5. *Dashboard & Deployment (Opsional)*

```
[ ] import matplotlib.pyplot as plt #visualisasi data
    import seaborn as sns #visualisasi data
```

*Gambar 9. Sintaks Visualisasi Data*

Kode tersebut mengimpor dua pustaka yang sering digunakan untuk visualisasi data di Python, yaitu matplotlib dan seaborn. matplotlib.pyplot adalah sub-modul dari pustaka matplotlib yang menyediakan berbagai fungsi dan metode untuk membuat plot grafik. Fungsi-fungsi ini mencakup pembuatan scatter plot, line plot, bar plot, histogram, dan banyak lagi. Dengan memberikan alias plt, kita dapat menggunakan fungsi-fungsi dari matplotlib.pyplot dengan lebih mudah. Di sisi lain, seaborn adalah pustaka Python yang memungkinkan pembuatan grafik yang lebih menarik secara estetika dan menyediakan berbagai fungsi untuk memplot berbagai jenis grafik dengan mudah. Fungsi-fungsi yang tersedia di seaborn

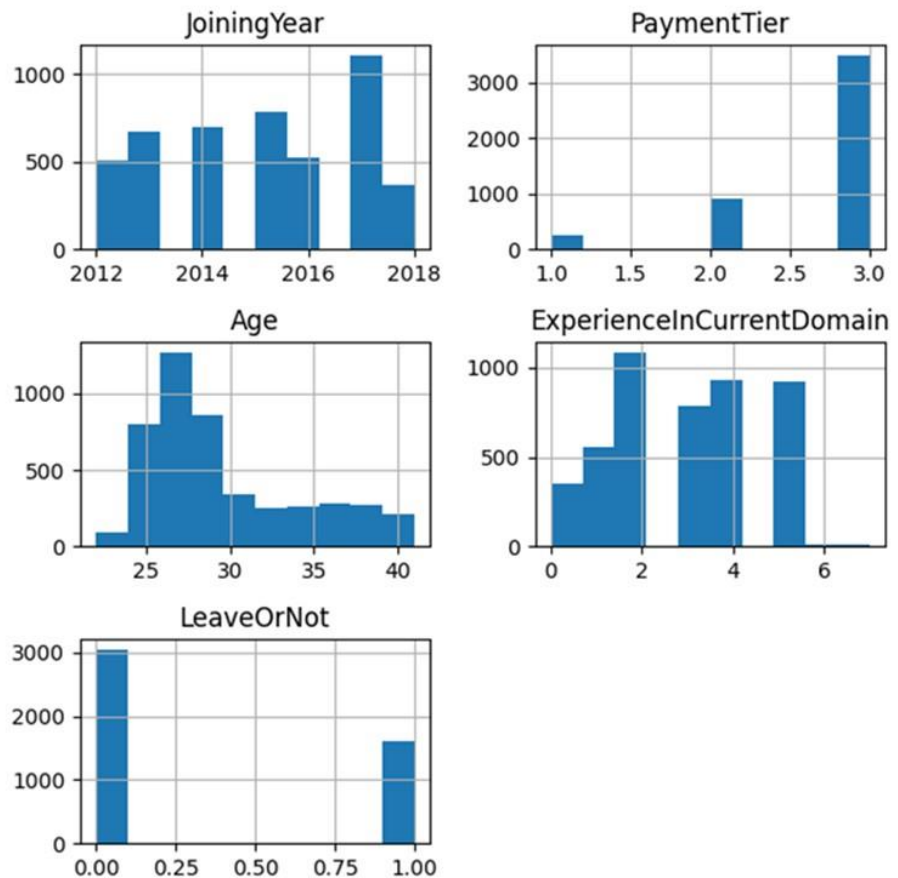
termasuk pembuatan scatter plot, bar plot, line plot, heatmap, dan lainnya. Dengan memberikan alias `sns`, kita dapat menggunakan fungsi-fungsi dari pustaka `seaborn` dengan lebih singkat dan intuitif. Kombinasi dari `matplotlib` dan `seaborn` memberikan fleksibilitas dan kemudahan dalam membuat visualisasi data yang informatif dan menarik.

```
[ ] df.hist(figsize=(6,6), bins=10)
    plt.tight_layout()
```

*Gambar 10. Sintaks Visualisasi Data*

Kode `df.hist(figsize=(6,6), bins=10)` digunakan untuk membuat histogram dari data yang terdapat dalam `DataFrame` `df`. Argumen `figsize=(6,6)` digunakan untuk menentukan ukuran gambar plot histogram (6 inci x 6 inci), sedangkan argumen `bins=10` digunakan untuk mengatur jumlah bins (kelompok) yang digunakan dalam histogram. Selanjutnya, kode `plt.tight_layout()` digunakan untuk mengatur tata letak (layout) plot secara otomatis agar lebih rapi dan proporsional. Fungsi ini memastikan bahwa elemen-elemen plot seperti label sumbu dan judul dapat ditampilkan dengan baik tanpa tumpang tindih.

Hasil dari kode ini adalah plot histogram dari data dalam `DataFrame` `df` dengan ukuran gambar yang ditentukan (6 inci x 6 inci), menggunakan 10 bins untuk mengelompokkan data, dan tata letak plot yang rapi dan terstruktur. Hasil visualisasi data sebagai berikut



Gambar 11. Visualisasi Data

## 6. Pembuatan Presentasi Hasil *Final Project*

Setelah *project* selesai, dilakukan pembuatan dokumen presentasi yang memuat:

- Business Understanding*, termasuk studi kasus yang diangkat.
- Data Understanding*, termasuk fitur dan *behavioral pattern* yang ditemukan.
- Data Preparation*, termasuk apa saja yang dilakukan untuk mempersiapkan data & *feature engineering* yang dilakukan.
- Pilihan metode/algoritma (model).
- Alasan pemilihan metode/algoritma (model) dan komparasinya.

- f. Hasil dan tingkat akurasi model.
- g. Metriks evaluasi model dan alasan pemilihannya.
- h. Interpretasi dari model dan hasil evaluasi.
- i. *Insight* yang didapat dari dataset.
- j. Saran (rekomendasi) yang bisa diberikan untuk insight yang ditemukan.

## **7. Hambatan dan Penyelesaian**

Hambatan yang dialami terjadi pada model bagian *forecasting* dengan menggunakan 4 algoritma hampir tidak menemukan yang cocok karena hasil yang membuat kami bingung apakah harus mengganti model atau bagaimana.

Tetapi setelah diperiksa dan dicoba kembali, ternyata ada sedikit kesalahan dan ditemukan titik terang bahwa algoritma Gradient Boosting Classifier adalah pilihan terbaik dalam model Klasifikasi ini.

## **8. Tugas & Kontribusi Individu**

Setelah pembagian tim selesai dilanjutkan dengan pengerjaan *final project*, untuk yang pertama pemilihan dataset yang akan digunakan. Dari 6 *dataset*, masing-masing menganalisis satu *dataset* dan saya mendapatkan bagian untuk menganalisis *dataset* 1 yang pada akhirnya dipilih sebagai *dataset* yang digunakan untuk *final project* Kelompok 3 – Mentor Asfaar

Selanjutnya melakukan *Data Preparation* dengan *cleaning* data, saya mengerjakan data transaksi level 3 untuk *handling missing value* dan *duplicate value*. Pada bagian *data visualization*, saya membantu rekan tim dalam memvisualisasikan data yang sudah selesai dimodelkan dan juga membuat *background* untuk *dashboard* di Google Collab juga.



Dan pada *Selection to Showcase*, saya diamanahi oleh satu tim untuk menjadi presenter utama pada hari pelaksanaan. Mempresentasikan hasil *final project* selama 7 menit dan rekan tim membantu dalam sesi tanya jawab selama 9 menit.

### c. Hasil Final Project

Judul *final project* kelompok 3 – Mentor Asfaar adalah “Analisis Regional Super App Memahami Pola Konsumsi Dan Preferensi Produk Di Beberapa Provinsi Indonesia”.

Model	Akurasi	Precision	Recall	F1-Score
KNeighbors	0.73	0.66	0.42	0.51
DTC	0.76	0.74	0.47	0.57
SVC	0.66	0.51	0.21	0.29
GradientBoost	0.79	0.84	0.47	0.61
KNeighbors (optimasi)	0.92	0.91	0.96	0.93
DTC (optimasi)	0.79	0.85	0.46	0.59
SVC (optimasi)	0.93	0.92	0.95	0.93
GradientBoost (optimasi)	0.93	0.94	0.96	0.95

### ***3.4.1 Sebelum Optimasi***

#### **3.4.1.1 Akurasi**

Berdasarkan tabel di atas, model gradient boost mempunyai akurasi yang terbaik yaitu 0.79. Artinya adalah 79% dari semua prediksi yang dibuat oleh model Gradientboost ini adalah benar.

#### **3.4.1.2 Presisi**

Berdasarkan tabel di atas, model gradient boost mempunyai presisi yang terbaik yaitu 0.84. Artinya adalah 84% dari data yang diprediksi model termasuk ke dalam kelas 1, ternyata adalah benar.

#### **3.4.1.3 Recall**

Berdasarkan tabel di atas, model DTC dan Gradient Boost mempunyai recall yang terbaik yaitu 0.47. Artinya Model DTC dan Gradient Boost mengidentifikasi dengan benar 47% dari data poin yang sebenarnya milik kelas 1.

#### **3.4.1.4 F-1 Score**

Berdasarkan tabel di atas, model Gradient Boost mempunyai F-1 Score yang terbaik yaitu 0.61. Artinya Model Gradient mempunyai ukuran yang seimbang antara presisi dan recall lebih baik dari model yang lain.

### ***3.4.2 Setelah Optimasi***

#### **3.4.2.1 Akurasi**

Berdasarkan tabel di atas, model gradient boost dan SVC mempunyai akurasi yang terbaik yaitu 0.93. Artinya adalah 93% dari semua prediksi yang dibuat oleh model Gradientboost ini adalah benar.

#### **3.4.2.2 Presisi**

Berdasarkan tabel di atas, model gradient boost mempunyai presisi yang terbaik yaitu 0.94. Artinya adalah 94% dari data yang diprediksi model termasuk ke dalam kelas 1, ternyata adalah benar.

#### **3.4.2.3 Recall**

Berdasarkan tabel di atas, model Gradient Boost dan KNN mempunyai recall yang terbaik yaitu 0.96. Artinya Model DTC mengidentifikasi dengan benar 96% dari data poin yang sebenarnya milik kelas 1.

#### **3.4.2.4 F-1 Score**

Berdasarkan tabel di atas, model Gradient Boost mempunyai F-1 Score yang terbaik yaitu 0.95. Artinya Model Gradient mempunyai ukuran yang seimbang antara presisi dan recall lebih baik dari model yang lain.

## II. Aktivitas Bulanan

Bulan	Kegiatan
1	<p>1. Saya mengikuti mentoring dengan Kakak Mentor dan cycle meeting dengan Kakak Trainer. Setiap minggu diadakan kelas sebanyak 3 kali untuk membahas lebih dalam terkait modul dan learning video yang telah dipelajari. Selain itu setiap minggu juga diadakan kelas meeting sebanyak 1 kali untuk meningkatkan soft skill. Saya aktif berdiskusi dengan Kakak Mentor dan teman teman dalam grup whatsapp. Setelah pengumuman pembagian DPP, saya diminta untuk masuk ke dalam grup whatsapp DPP. Saya dan teman-teman dapat berkonsultasi terkait konversi sks dan mata kuliah yang sekiranya cocok untuk dikonversikan.</p> <p>2. Saya telah mempelajari modul dan record zoom mengulang video-video supaya lebih mengerti dan jelas serta mengerjakan kuis dari setiap materi yang telah trainer jelaskan. saat ini saya telah mempelajari materi mengenai SQL dan Datawarehouse ETL. Saya sangat senang dan tertantang walaupun banyak error ketika mencoba tutor tutor dari trainer, tetapi disitulah pengalaman saya semakin bertambah</p> <p>3. Tantangan yang saya hadapi di bulan ini adalah mengerjakan pre test dan post test 1. Pre test merupakan ujian dari Celerates untuk mengetahui kemampuan awal seluruh peserta sebelum mempelajari modul Data Science di Celerates. Saya menyelesaikan pre test dengan belajar terlebih dahulu sebelumpelaksanaan ujian. Selain itu terdapat post test yang merupakan test individu setelah mempelajari modul modul yang diberikan, ini untuk melatih kemampuan kepada semua peserta seberapa paham modul yang telah dipahami. Sebelum mengerjakan post test biasanya saya akan bertanya di grup whatsapp dan bertanya pada saat mentoring untuk menambah pemahaman dan mempersiapkan sebelum post test.</p> <p>4. Pengembangan kompetensi yang telah saya dapatkan pada bulan ini adalah mempelajari SQL dengan tools menggunakan postgre SQL dan dbriver. Saya dapat memahami informasi umum tentang data science, cara mengolah data, beberapa teori dasar, dan pentingnya data science. Saya juga memahami Datawarehouse. Saya juga belajar memiliki kemampuan berpikir statistik yang dapat diterapkan untuk memahami data dalam konteks bisnis. Selain pemahaman tentang SQL dan</p>

	<p>Datawarehouse, saya mempelajari cara membangun dan menjaga semangat belajar serta kepemimpinan pada diri sendiri. Saya juga belajar untuk mengetahui cara mempersiapkan diri agar siap bekerja.</p>
--	--

2	<p>1. Saya mengikuti mentoring dengan Kakak Mentor dan cycle meeting dengan Kakak Trainer. Setiap minggu diadakan kelas sebanyak 3 kali untuk membahas lebih dalam terkait modul dan learning video yang telah dipelajari. Di bulan ini saya belajar modul mengenai basic basic bahasa pemrograman python. Selain itu setiap minggu juga diadakan kelas meeting sebanyak 1 kali untuk meningkatkan soft skill. Saya aktif berdiskusi dengan Kakak Mentor dan teman teman dalam grup whatsapp. DPP saya selalu memonitoring terhadap mahasiswa yang dibimbingnya melalui grup whatsapp. Selalu menanyakan bagaimana pembelajarannya dan sangat terbuka untuk berkonsultasi terkait konversi sks dan masalah selama pembelajaran yang sedang berjalan.</p> <p>2. Saya telah mempelajari modul dan record zoom mengulang video-video supaya lebih mengerti dan jelas serta mengerjakan kuis dari setiap materi yang telah trainer jelaskan. saat ini saya telah mempelajari materi mengenai Basic basic Bahasa Pemrograman Python. Saya sangat senang dan tertantang walaupun banyak error ketika mencoba tutor tutor dari trainer, tetapi disitulah pengalaman saya semakin bertambah</p> <p>3. Tantangan yang saya hadapi di bulan ini adalah mengantuk dan lemas ketika belajar dikarenakan menahan haus dan lapar saat bulan ramadhan sehingga materi yang saya tonton dari zoom itu tidak maksimal saya pahami, tetapi untungnya celerates menyediakan record zoom nya sehingga saya bisa menonton berulang kali video dari trainer nya sampai paham. Dan mungkin tantangan selanjutnya yaitu di mengerjakan pre test dan post test, Pre test merupakan ujian dari Celerates untuk mengetahui kemampuan awal seluruh peserta sebelum mempelajari modul Python di Celerates. Saya menyelesaikan pre test dengan belajar terlebih dahulu sebelum pelaksanaan ujian. Selain itu terdapat post test yang merupakan test individu setelah mempelajari modul modul yang diberikan, ini untuk melatih kemampuan kepada semua mahasiswa seberapa paham modul yang telah dipelajari. Sebelum mengerjakan post test biasanya saya akan bertanya di grup whatsapp dan bertanya pada saat mentoring untuk menambah pemahaman dan mempersiapkan sebelum post test.</p> <p>4. Pengembangan kompetensi yang telah saya dapatkan pada bulan ini adalah</p>
---	--

	<p>menambah ilmu basic dari bahasa pemrograman Python menggunakan Google Collabs. Saya dapat memahami tentang Tipe data, TUPLE, Dictionary, If else, for, block, While, Exception, Function statement. Saya juga memahami mengenai Lambda, Class, Numpy, Pandas. Dan yang terakhir saya belajar mengenai</p> <ol style="list-style-type: none"> <li>1. Create Data</li> <li>2. Get Data</li> <li>3. Prepare Data</li> <li>4. Analyze</li> <li>5. Present Data</li> </ol> <p>saya juga mempelajari cara membangun dan menjaga semangat belajar serta kepemimpinan pada diri sendiri. Saya juga belajar untuk mengetahui cara mempersiapkan diri agar siap bekerja.</p>
--	---

3	<p>1. Saya mengikuti mentoring dengan Kakak Mentor dan cycle meeting dengan Kakak Trainer. Setiap minggu diadakan kelas sebanyak 3 kali untuk membahas lebih dalam terkait modul dan learning video yang telah dipelajari. Di bulan ini saya belajar modul mengenai Machine Learning.</p> <p>Selain itu kadang setiap minggu juga diadakan kelas meeting sebanyak 1 kali untuk meningkatkan soft skill. Saya aktif berdiskusi dengan Kakak Mentor dan teman teman dalam grup whatsapp.</p> <p>DPP saya selalu memonitoring terhadap mahasiswa yang dibimbingnya melalui grup whatsapp. Selalu menanyakan bagaimana pembelajarannya dan sangat terbuka untuk berkonsultasi terkait konversi sks dan masalah selama pembelajaran yang sedang berjalan.</p> <p>2. Saya telah mempelajari modul Machine Learning dan record zoom mengulang video-video supaya lebih mengerti dan jelas serta mengerjakan kuis dari setiap materi yang telah trainer jelaskan. saat ini saya telah mempelajari materi mengenai Machine Learning dan Library nya yaitu terdapat Pandas, Numpy, matplotlib dan lain lain. Saya sangat senang dan tertantang walaupun banyak error ketika mencoba tutor tutor dari trainer, tetapi disitulah pengalaman saya semakin bertambah</p> <p>3. Tantangan yang saya hadapi di bulan ini adalah kurang paham terhadap modul Machine Learning dikarenakan tidak mendapatkan ilmu ilmu basic machine learning dan sepertinya saya agak kaget dengan modul seperti ini, tetapi untungnya celerates menyediakan record zoom nya sehingga saya bisa menonton berulang kali video dari trainer nya sampai paham. Dan mungkin tantangan selanjutnya yaitu di mengerjakan pre test dan post test, Pre test merupakan ujian dari Celerates untuk mengetahui kemampuan awal seluruh peserta sebelum mempelajari modul Phyton di Celerates.</p> <p>Saya menyelesaikan pre test dengan belajar terlebih dahulu sebelum pelaksanaan ujian. Selain itu terdapat post test yang merupakan test individu setelah mempelajari modul modul yang diberikan, ini untuk melatih kemampuan kepada semua mahasiswa seberapa paham modul yang telah dipelajari. Sebelum mengerjakan post test biasanya saya akan bertanya di grup whatsapp dan bertanya pada</p>
---	--



	<p>saat mentoring untuk menambah pemahaman dan mempersiapkan sebelum post test.</p> <p>4. Pengembangan kompetensi yang telah saya dapatkan pada bulan ini adalah menambah ilmu ilmu terkait Machine Learning dan Deep Learning menggunakan Google Collabs. Saya dapat memahami tentang Data Preprocessing, Data Cleaning, Memodelkan Data, Memvisualisasikan Data. Saya juga memahami mengenai terkait Deep Learning walaupun baru setengah perjalanan. saya juga mempelajari cara membangun dan menjaga semangat belajar serta kepemimpinan pada diri sendiri. Saya juga belajar untuk mengetahui cara mempersiapkan diri agar siap bekerja.</p>
--	---

4	<p>1. Saya mengikuti mentoring dengan Kakak Mentor dan cycle meeting dengan Kakak Trainer. Setiap minggu diadakan kelas sebanyak 3 kali untuk membahas lebih dalam terkait modul dan learning video yang telah dipelajari. Di bulan ini saya belajar modul mengenai Deep Learning. Selain itu kadang setiap minggu juga diadakan kelas meeting sebanyak 1 kali untuk meningkatkan soft skill. Saya aktif berdiskusi dengan Kakak Mentor dan teman teman dalam grup whatsapp. DPP saya selalu memonitoring terhadap mahasiswa yang dibimbingnya melalui grup whatsapp. Selalu menanyakan bagaimana pembelajarannya dan sangat terbuka untuk berkonsultasi terkait konversi sks dan masalah selama pembelajaran yang sedang berjalan.</p> <p>2. Saya telah mempelajari modul Deep Learning dan record zoom mengulang video-video supaya lebih mengerti dan jelas serta mengerjakan kuis dari setiap materi yang telah trainer jelaskan. saat ini saya telah mempelajari materi mengenai Deep Learning dan Library nya yaitu terdapat Pandas, Numpy, matplotlib dan lain lain. Saya sangat senang dan tertantang walaupun banyak error ketika mencoba tutor tutor dari trainer, tetapi disitulah pengalaman saya semakin bertambah.</p> <p>3. Tantangan yang saya hadapi di bulan ini adalah kurang paham terhadap modul Deep Learning dikarenakan tidak mendapatkan ilmu ilmu basic Deep learning dan sepertinya saya agak kaget dengan modul seperti ini, tetapi untungnya celerates menyediakan record zoom nya sehingga saya bisa menonton berulang kali video dari trainer nya sampai paham. Dan mungkin tantangan selanjutnya yaitu di mengerjakan pre test dan post test, Pre test merupakan ujian dari Celerates untuk mengetahui kemampuan awal seluruh peserta sebelum mempelajari modul Phyton di Celerates. Saya menyelesaikan pre test dengan belajar terlebih dahulu sebelum pelaksanaan ujian. Selain itu terdapat post test yang merupakan test individu setelah mempelajari modul modul yang diberikan, ini untuk melatih kemampuan kepada semua mahasiswa seberapa paham modul yang telah dipelajari. Sebelum mengerjakan post test biasanya saya akan bertanya di grup whatsapp dan bertanya pada saat mentoring untuk menambah pemahaman dan mempersiapkan sebelum post test.</p> <p>4. Pengembangan kompetensi yang telah saya dapatkan pada bulan ini adalah menambah ilmu ilmu terkait Deep Learning menggunakan Google Collabs. Saya dapat memahami tentang</p>
---	--

	<p>Data Preprocessing, Data Cleaning, Memodelkan Data, Memvisualisasikan Data. Saya juga memahami mengenai terkait Deep Learning walaupun baru setengah perjalanan. saya juga mempelajari cara membangun dan menjaga semangat belajar serta kepemimpinan pada diri sendiri. Saya juga belajar untuk mengetahui cara mempersiapkan diri agar siap bekerja.</p>
--	---

5	<p>1. Saya mengikuti mentoring dengan Kakak Mentor dan cycle meeting dengan Kakak Trainer. Setiap minggu diadakan kelas sebanyak 3 kali untuk membahas lebih dalam terkait modul dan learning video yang telah dipelajari. Di bulan ini saya sedang menggarap final projek, dan saya masuk dalam sebuah kelompok yang beranggotakan 5 orang. Selain itu kadang setiap minggu juga diadakannya kelas meeting sebanyak 1 kali untuk meningkatkan soft skill. Saya aktif berdiskusi dengan Kakak Mentor dan teman teman dalam grup whatsapp. DPP saya selalu memonitoring terhadap mahasiswa yang dibimbingnya melalui grup whatsapp. Selalu menanyakan bagaimana pembelajarannya dan sangat terbuka untuk berkonsultasi terkait konversi sks dan masalah selama pembelajaran yang sedang berjalan.</p> <p>2. Saya mengulang modul Machine Learning dan record zoom mengulang video-video supaya lebih mengerti dan jelas serta mengerjakan kuis dari setiap materi yang telah trainer jelaskan. saat ini saya ini saya sedang menggarap final projek yang berjudul "EMPLOYEE CHURN RATE PREDICTION LEAVE OR NOT". Dan kelompok kami menggunakan metode Classification Saya sangat senang dan tertantang walaupun banyak error ketika mencoba tutor tutor dari trainer, tetapi disitulah pengalaman saya semakin bertambah.</p> <p>3. Tantangan yang saya hadapi di bulan ini adalah kurang paham terhadap sintaks sintaks di modul machine learning dan terhadap pemodelan dan optimasi model yang tepat itu seperti apa. Mungkin tantangan selanjutnya yaitu di pembagian tugas nya yang sedikit kurang efektif karena pengerjaan nya secara online tetapi permasalahan tersebut bisa pecahkan dengan selalu berkomunikasi terus menerus di grup WhatsApp dan menggunakan tools yang bisa dikerjakan bareng bareng.</p> <p>4. Pengembangan kompetensi yang telah saya dapatkan pada bulan ini adalah menambah ilmu ilmu terkait Machine Learning menggunakan Google Collabs. Saya dapat memahami tentang Data Preprocessing, Data Cleaning, Memodelkan Data, Memvisualisasikan Data. Saya juga memahami mengenai Model model data yaitu seperti KneighborClassifier, Gradient Boosting, Support Vector Machine (SVM) dll. Dan saya juga dapat memahami cara menganalisis hasil yang baik dan benar.saya juga</p>
---	---

	<p>mempelajari cara membangun dan menjaga semangat belajar serta kepemimpinan pada diri sendiri. Pada final proyek ini Saya mendapatkan softskill bagaimana cara bekerja sama, komunikasi, dan saling support bersama dengan tim.</p>
--	---

#### **A. Sertifikasi Keahlian**

Sertifikasi : Kelulusan Celerates - Data Science Basic

Tanggal Pelaksanaan : 16 Februari - 30 Juni 2024

### **III. Penutup**

#### **A. Kesimpulan**

Program Celerates Batch 6 - Data Science Basic adalah studi independen bersertifikat yang diselenggarakan oleh PT Mitra Talenta Grup dengan dukungan Kemdikbud Ristekdikti. Program ini berlangsung dari Februari hingga Juni 2024, dirancang sepenuhnya secara daring untuk mempersiapkan talenta digital di bidang data science yang kompetitif di era digital.

#### **1. Lingkup Program**

##### **a. Deskripsi Kegiatan:**

- Materi Pelatihan: Disampaikan secara daring mencakup SQL, Python, statistika, machine learning, deep learning dan visualisasi data
- Mentorship dan Networking: Peserta mendapat pendampingan intensif dari mentor profesional serta kesempatan networking dengan pemain industri digital.
- Proyek Kolaboratif: Peserta menyelesaikan proyek bersama untuk pengalaman nyata.

##### **b. Tujuan Program:**

- Mengembangkan kompetensi peserta dalam data science.
- Memampukan peserta berkolaborasi dan mengerjakan proyek nyata.
- Menyiapkan peserta untuk berkarier di industri digital.

##### **c. Kurikulum Program:**

- Foundation for Data Science: Mengenalkan dasar-dasar data science, SQL, Python, dan statistika.
- Data Preprocessing: Fokus pada persiapan dan pengolahan data.
- Machine Learning: Pembelajaran model dan algoritma machine learning serta evaluasinya.

- Data Visualization: Visualisasi data menggunakan BI Tools dan penyampaian hasil analisis.
- d. Metode Pembelajaran:
- Asinkronus: Pembelajaran mandiri melalui modul, video, kuis, dan referensi lain di Learning Management System.
  - Sinkronus: Kelas interaktif, diskusi, dan sesi mentoring dengan expert.
  - Penilaian: Tugas individu/kelompok, keaktifan, kuis, dan proyek akhir.
- e. Penilaian:
- Tugas Individu: Membaca referensi dan kuis.
  - Keaktifan: Interaksi selama sesi sinkronus.
  - Mentorship: Pendampingan dari mentor dan fasilitator.
  - Proyek Tim: Penyelesaian proyek akhir yang relevan dengan industri.
  - Showcase: Presentasi proyek final di depan juri.

Program ini menyediakan pembelajaran mendalam dan praktis dalam data science, dengan dukungan penuh melalui mentorship dan networking. Dengan kurikulum yang dirancang untuk memenuhi kebutuhan industri, program ini bertujuan untuk mencetak talenta digital yang siap berkontribusi di dunia kerja. Proses pembelajaran yang komprehensif dan sistem penilaian yang beragam memastikan peserta memperoleh kompetensi yang diperlukan untuk berkarier di industri digital.

## **B. Saran**

Berdasarkan deskripsi program Celerates Batch 6 - Data Science, berikut beberapa saran untuk meningkatkan efektivitas dan kualitas program:

1. Penambahan Konten Praktis dan Studi Kasus Industri:

- **Praktik Nyata:** Tambahkan lebih banyak studi kasus dari industri untuk memberikan gambaran nyata kepada peserta tentang bagaimana data science diterapkan dalam dunia kerja.
- **Kolaborasi dengan Perusahaan:** Ajak perusahaan untuk memberikan proyek nyata yang dapat dikerjakan oleh peserta sebagai bagian dari kurikulum.

## 2. Penilaian dan Umpan Balik yang Komprehensif:

- **Feedback Berkelanjutan** Berikan umpan balik yang terperinci dan berkelanjutan pada setiap tugas dan proyek, sehingga peserta dapat terus memperbaiki kemampuan mereka.
- **Penilaian Peer-to-Peer:** Libatkan peserta dalam penilaian peer-to-peer untuk meningkatkan keterlibatan dan tanggung jawab.

## 3. Penggunaan Teknologi Terbaru:

- **Perkembangan Teknologi** Pastikan kurikulum terus diperbarui sesuai dengan perkembangan terbaru dalam teknologi data science
- **Alat dan Platform Modern:** Kenalkan peserta pada alat dan platform terbaru yang digunakan di industri, seperti software machine learning terbaru atau layanan cloud yang canggih.

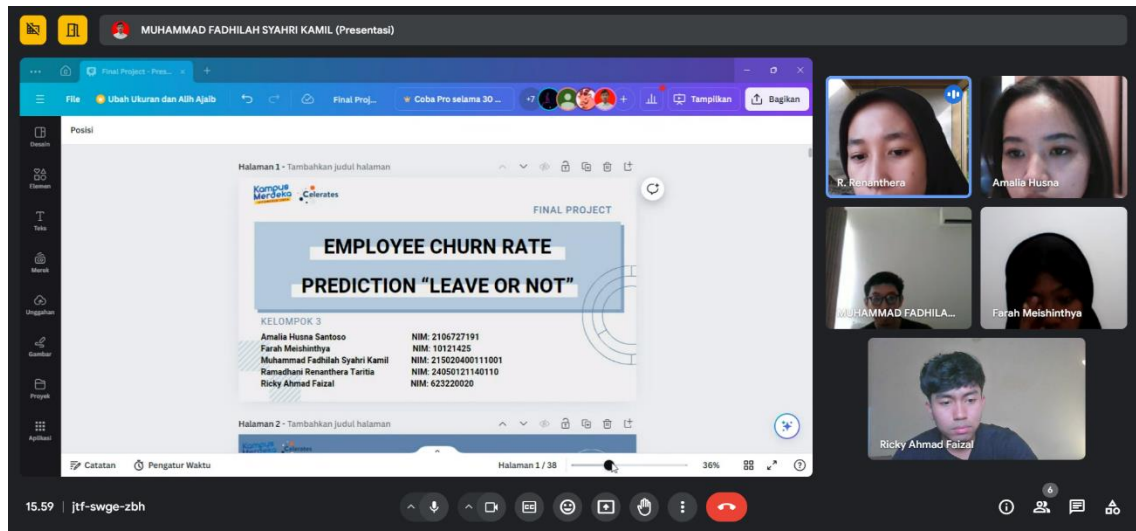


## Referensi

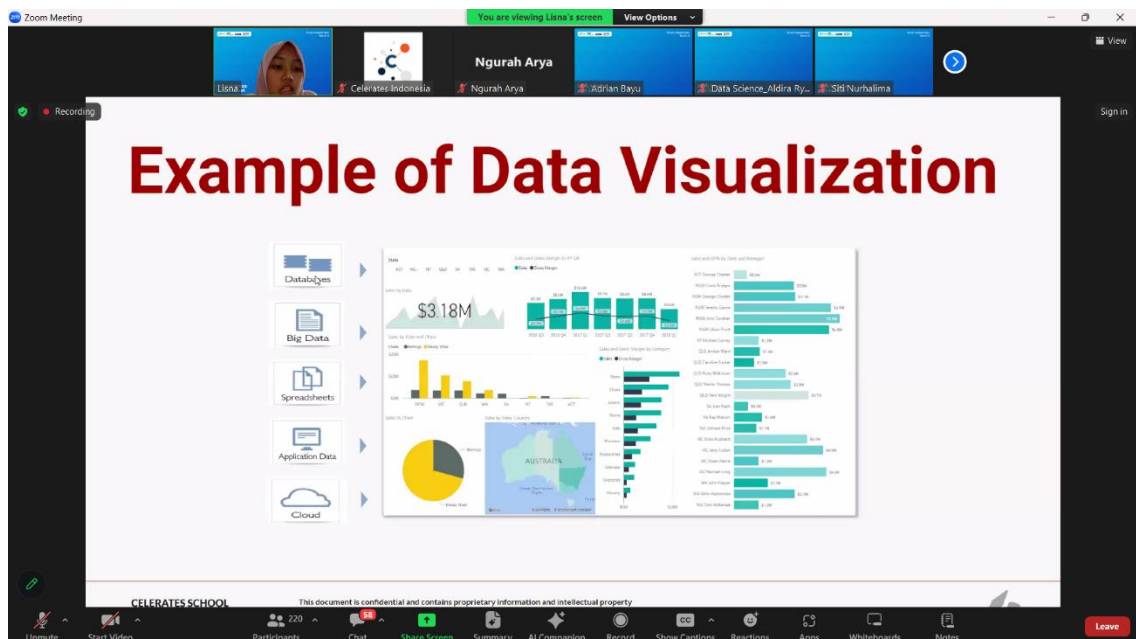
- [1] [Online]. Available: <https://kampusmerdeka.kemdikbud.go.id/>.
- [2] S. D. Sendjaja, Pengantar Teori Komunikasi, Universitas Terbuka, 2014.
- [3] Startup Campus, “GUIDELINE Final Project Data Science & AI Startup Campus Batch 6,” [Online]. Available: [https://drive.google.com/file/d/1rNf\\_NBSods4WdQMZIEjbVfRgPSAuMQ55/view](https://drive.google.com/file/d/1rNf_NBSods4WdQMZIEjbVfRgPSAuMQ55/view).
- [4] “Tableau,” [Online]. Available: <https://www.coursera.org/articles/tableau>.
- [5] “Forecasting,” [Online]. Available: <https://accurate.id/bisnis-ukm/apa-ituforecasting/>.
- [6] “Wikipedia,” [Online]. Available: [https://en.wikipedia.org/wiki/Feature\\_scaling](https://en.wikipedia.org/wiki/Feature_scaling).
- [7] “Clustering,” [Online]. Available: <https://www.coursera.org/articles/clustering>.

## Lampiran

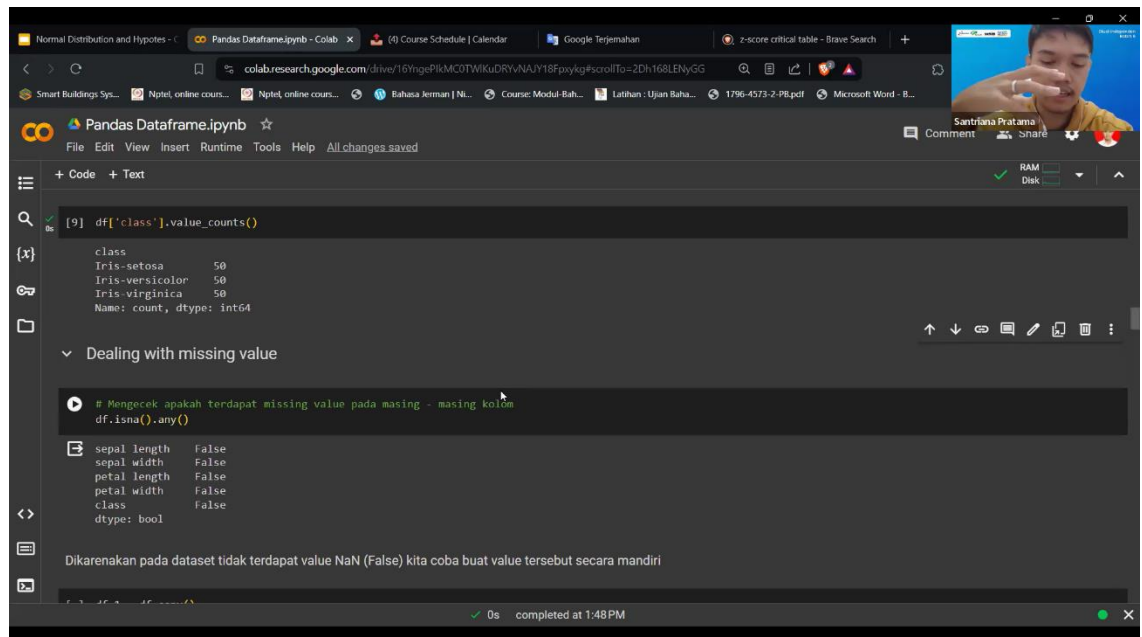
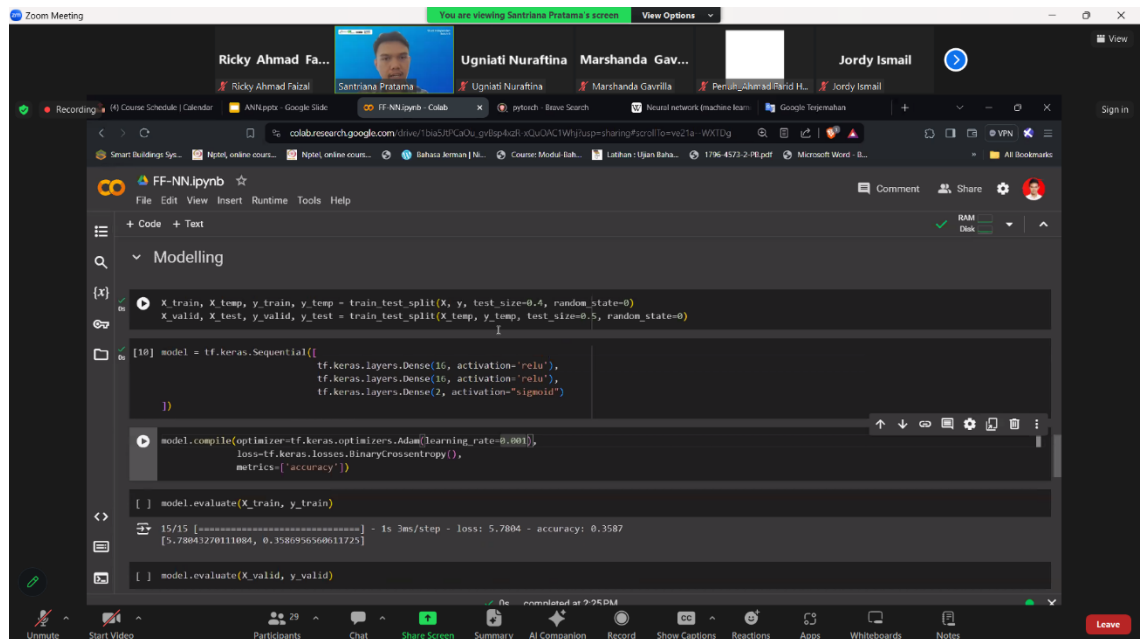
### 1. Dokumentasi *Final Project*



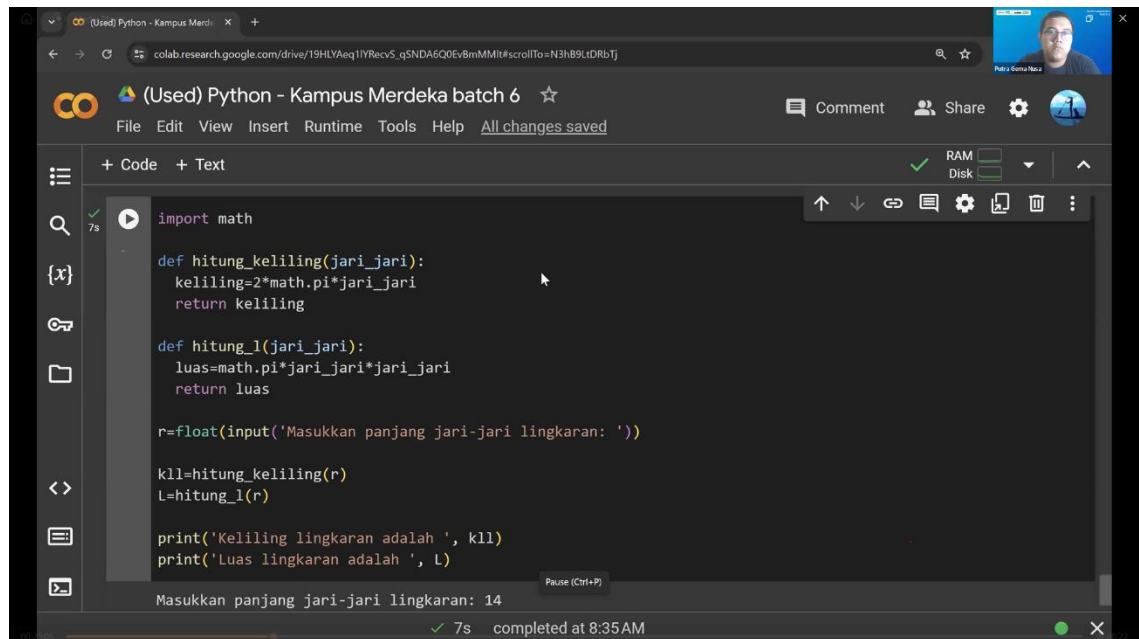
### 2. Dokumentasi *Kelas Synchronus*



### 3. Dokumentasi *Kelas Marticular*



#### 4. Dokumentasi Mentoring



```
import math

def hitung_keliling(jari_jari):
    keliling=2*math.pi*jari_jari
    return keliling

def hitung_l(jari_jari):
    luas=math.pi*jari_jari*jari_jari
    return luas

r=float(input('Masukkan panjang jari-jari lingkaran: '))

kll=hitung_keliling(r)
L=hitung_l(r)

print('Keliling lingkaran adalah ', kll)
print('Luas lingkaran adalah ', L)
```

Masukkan panjang jari-jari lingkaran: 14

7s completed at 8:35AM

## 5. Kelas Softskill



Kampus Merdeka MSIB Celerates

Studi Independen Celerates Batch 6 Proudly Present: Softskill Class

# Basic Personal Branding POV Self Development

Speakers:

  
Meisya Sallwa  
Penulis Buku & Public Speaker

## 6. Dokumentasi Pembekalan Final Project

Final Project Data Science & A

Course Schedule | Calendar

C:\Users\UISNA%20SULSTIAWATI\Downloads\Final%20Project%20Data%20Science%20&%20Analyst.docx(1).pdf

Final Project Data Science & Analyst.docx

1 / 21 | - 100% + |

Balance  
Lisaka  
gaji ya

# Final Project Kampus Merdeka

Program : Data Science & Analyst  
Periode : 03 Juni - 21 Juni 2024

Celerates

INDONESIA

PT Mitra Talenta Grup

Instagram : @celerates.id & @celerates.school

website : [www.celerates.co.id](http://www.celerates.co.id)