

LAPORAN AKHIR
MAGANG & STUDI INDEPENDEN BERSERTIFIKAT
CELERATES ACCELERATION PROGRAM (BIG DATA & BUSINESS
INTELLIGENCE)
PT MITRA TALENTA GRUP (CELERATES)

Diajukan untuk memenuhi persyaratan kelulusan
Program MSIB MBKM

oleh:

Wahyu Citra Pratama / 20190801275



TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS ESA UNGGUL

2022

Nomor IA : 569/IA/BPKI-KS/TI/UEU/2022

**LEMBAR PENGESAHAN PROGRAM STUDI TEKNIK INFORMATIKA
UNIVERSITAS ESA UNGGUL**

BIG DATA & BUSINESS INTELIGENCE

Di PT. MITRA TALENTA GRUP (CELERATES)

oleh:

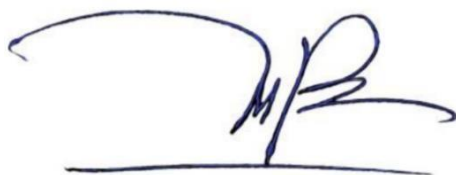
Wahyu Citra Pratama / 20190801275

disetujui dan disahkan sebagai

Laporan Magang atau Studi Independen Bersertifikat Kampus Merdeka

Bekasi, 25 Desember 2022

Pembimbing Magang atau Studi Independen Teknik Informatika Universitas Esa
Unggul



Ir. Nixon Erzed, MT.

NIDN: 0315116501

LEMBAR PENGESAHAN

BIG DATA & BUSINESS INTELIGENCE

Di PT. MITRA TALENTA GRUP (CELERATES)

oleh :

Wahyu Citra Pratama / 20190801275

disetujui dan disahkan sebagai

Laporan Magang atau Studi Independen Bersertifikat Kampus Merdeka

Bekasi, 25 December 2022

Pembimbing Lapangan,



Nina Retna Utami

NIP.MTG-TF/2018100029

Abstraksi

MSIB merupakan bagian dari program Kampus Merdeka yang bertujuan memberikan kesempatan pada mahasiswa untuk belajar dan mengembangkan diri melalui aktivitas di luar kelas perkuliahan. Terdapat banyak mitra yang ikut berkontribusi dalam mengasah pengetahuan mahasiswa salah satunya PT.Mitra Talenta Grup . Program MSIB Basis Data dan Business Intelligence merupakan bootcamp singkat Celerates dan pengerjaan final project bersertifikasi yang memiliki program yang berbeda , salah satunya adalah Big Data Dalam program ini, ada 9 teori utama yang akan diajarkan yaitu Database & SQL, Python for Data Analytics, Data Warehouse, Etl(Pentaho), Statistix, Data Visualization(Tableau & Power BI), Big Data & Cloud Technology, Machine Learning, Deep Learning. Diharapkan setelah mengikuti program ini, peserta menguasai dasar dari studi sebuah data khususnya Big Data & Business Intelligence dan dapat mengaplikasikannya pada final project berupa pembuatan Studi basis data dari persiapan data, membangun model, memprediksi kumpulan data dan membuat hasil akhir dalam bentuk dashboard. Mahasiswa/I yang lulus pada program ini diharapkan mampu membuat sebuah data analisis, menyajikan data yang dibuat oleh perusahaan dan mampu berkerja sebagai Data analyst, Data Scientist dan lainnya.

Kata Kunci : PT.Mitra Talenta Grup,SIB,Kampus Merdeka

Kata Pengantar

Segala puji bagi Allah SWT, Tuhan semesta alam yang telah memberikan taufik dan hidayah-Nya untuk kita semua. Shalawat serta salam selalu tercurah kepada junjungan kita Nabi Muhammad SAW yang kita nantikan syafaatnya di akhirat kelak. Dalam kesempatan kali ini, penulis ingin menyampaikan ucapan terima kasih kepada semua pihak yang telah membantu serta membimbing penulis dalam melaksanakan salah satu program kampus merdeka yakni Magang atau Studi Independen Bersertifikat 2022 sampai dengan tersusunnya laporan ini, kepada:

1. Ir. Nixon Erzed, MT. selaku Direcktur operasi kampus Bekasi Dan Dosen Pembimbing yang telah memberikan izin melaksanakan Magang Studi Independen Bersertifikat di di PT. Mitra Talenta Grup (Big Data & Business Intelligence)
2. Muhammad Bahrul Ulum, S.ko, M.Kom, MT.selaku Ketua Prodi Teknik Informatika yang telah memberikan izin melaksanakan Magang Studi Independen Bersertifikat di PT. Mitra Talenta Grup (Big Data & Business Intelligence)
3. Mba Nina Retno Utami selaku Mentor kelas Celerates Big Data
4. Mas Wildan Selaku Mentor Kelas Celerates Big Data
5. Mas Asfar Jaya selaku Mentor Kelompok kelas Celerates
6. Seluruh pengurus, dan mentor PT. Mitra Talenta Grup (Big Data & Business Intelligence) lainnya yang sudah membimbing saya dalam menyelesaikan kegiatan Studi Independen ini.

Akhir kata, semoga laporan Magang atau Studi Independen Bersertifikat ini dapat berguna, baik bagi penulis ataupun bagi pembaca. Tersadar dari itu semua, penulis menyadari bahwa dalam penyusunan laporan ini masih jauh dari kata sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran dari pembaca.

Bekasi, 25 Desember 2022



Wahyu Citra Pratama

Daftar Isi

Lembar Pengesahan Program Studi Teknik Informatika Universitas Esa Unggul..	i
Lembar Pengesahan	ii
Abstraksi	iii
Kata Pengantar	iv
Daftar Isi	v
Daftar Gambar	vii
BAB I Pendahuluan	8
1.1. Latar belakang	8
1.2. Lingkup Kerja	8
1.3. Tujuan.....	8
BAB II Organisasi atau Lingkungan PT. Mitra Talenta Grup	10
2.1. Struktur Organisasi.....	10
2.2. Lingkup Pekerjaan.....	10
2.3. Deskripsi Pekerjaan	10
2.4. Jadwal Kerja	11
BAB III Celerates Acceleration Program (Big Data & Business Intelligence)	12
3.1. Deskripsi Projek Akhir	12
3.2. Metode Penelitian.....	12
3.3. Proses Pelaksanaan Projek	13
3.3.1. Analisis Data	13
3.3.2. Pengolahan Data	17
3.4. Hasil Pencapaian Projek	20

3.4.1. Visualisasi Data.....	20
BAB IV Penutup	23
4.1. Kesimpulan.....	23
4.2. Saran.....	23
BAB V Lampiran A. TOR	A - 1
BAB VI Lampiran B. Log Activity	B - 1
BAB VII Lampiran C. Dokumen Teknik.....	C - 1
BAB VIII Lampiran D. Surat Pernyataan Kerahasiaan	D - 1

Daftar Gambar

Gambar 1 Analisis Data 1.....	13
Gambar 2 Analisis Data 2.....	13
Gambar 3 Analisis Data 3.....	14
Gambar 4 Analisis Data 4.....	14
Gambar 5 Analisis Data 5.....	15
Gambar 6 Analisis Data 6.....	15
Gambar 7 Analisis Data 7.....	15
Gambar 8 Analisis Data 8.....	16
Gambar 9 Analisis Data 9.....	16
Gambar 10 Analisis Data 10.....	16
Gambar 11 Analisis Data 11.....	17
Gambar 12 Analisis Data 12.....	17
Gambar 13 Pengolahan Data 1	17
Gambar 14 Pengolahan Data 2	18
Gambar 15 Pengolahan Data 3	18
Gambar 16 Pengolahan Data 4	19
Gambar 17 Pengolahan Data 5	19
Gambar 18 Pengolahan Data 6	19
Gambar 19 Pengolahan Data 7	20
Gambar 20 Pengolahan Data 8	20
Gambar 21 Visualisasi Data 1	21
Gambar 22 Visualisasi Data 2	21

BAB I

Pendahuluan

1.1. Latar belakang

Kampus Merdeka merupakan bagian dari kebijakan Merdeka Belajar oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia yang memberikan kesempatan bagi mahasiswa/i untuk mengasah kemampuan sesuai bakat dan minat dengan terjun langsung ke dunia kerja sebagai persiapan karier masa depan. Kampus merdeka tersebut merupakan kerangka untuk menyiapkan mahasiswa menjadi sarjana yang tangguh, relevan dengan kebutuhan zaman, dan siap menjadi pemimpin dengan semangat kebangsaan yang tinggi. Banyak program kampus merdeka yang dapat di ikuti mahasiswa guna mengasah skill untuk mempersiapkan karier di masa depan, salah satunya adalah program Studi Independen Bersertifikat, Studi Independen Bersertifikat adalah bagian dari program Kampus Merdeka yang bertujuan untuk memberikan kesempatan kepada mahasiswa untuk belajar dan mengembangkan diri melalui aktivitas di luar kelas perkuliahan, namun tetap diakui sebagai bagian dari perkuliahan.

PT. Mitra Talenta Grup bekerja sama dengan MSIB Kampus merdeka untuk melaksanakan studi independen Celerates Acceleration adalah sebuah sarana untuk mengembangkan kemampuan yang fokus pada System & Application Development dan Business Intelligence.

1.2. Lingkup Kerja

Sebagai Adapun lingkup projek selama mengikuti Magang dan Studi Independen Bersertifikat yaitu :

1. Pembekalan materi pelatihan seara daring
2. Pendampingan dan pembelajaran bermasama mentor dan trainer
3. Final Project Basis data dan kasus masalah
4. Adanya Quiz dan Pengerjaan materi dari pengerjaan tugas yang di kerjakan

1.3. Tujuan

Adapun tujuan dari kegiatan Magang Studi Independen Bersertifikat yaitu:

1. Mengasah kemampuan Problem Solving dan pengenalan data melalui praktik.

2. Menerapkan Business Intelligence menjelaskan tentang suatu konsep dan metode bagaimana untuk meningkatkan kualitas pengambilan keputusan bisnis berdasarkan sistem yang berbasis data
3. Mempersiapkan diri untuk berkarier di masa depan pada bidang Basis data dan business intelligence.

BAB II

Organisasi atau Lingkungan PT. Mitra Talenta Grup

PT Mitra Talenta Group (Celerates) adalah perusahaan yang bergerak dalam bidang pengembangan sumber daya manusia dalam bidang IT yang bertujuan untuk membantu talenta-talenta ini meng-akselerasi karir mereka. Celerates hadir sebagai perusahaan yang memberikan solusi IT bagi perusahaan yang ingin mengembangkan bisnis dengan teknologi yang terkini. Celerates juga menyediakan resource talent IT sesuai dengan kebutuhan perusahaan dan terus diberikan pelatihan untuk mengikuti trend teknologi yang selalu digunakan saat ini

2.1. Struktur Organisasi

Adapun struktur organisasi dari PT Mitra Talenta Group (Celerates) bersifat rahasia dan dilampirkan surat pernyataan kerahasiaan pada halaman Lampiran.

2.2. Lingkup Pekerjaan

Pada program Celerates Acceleration Program ini berfokus pada System & Application Development dan Business Intelligence. Modul pembelajarannya meliputi :

1. Database & SQL
2. Python for Data Analytics
3. ETL Pentaho
4. Statistic
5. Data Visualitization (Tableau & Power BI)
6. Big Data & Cloud Technology
7. Machine Learning
8. Deep Learning
9. Data Warehouse

2.3. Deskripsi Pekerjaan

Aktivitas Studi Independen Celerates Acceleration Program ini meliputi pembelajaran individu, dimana setiap peserta akan memiliki pembimbing sebagai tempat konsultasi jika ditemui kesulitan non-akademik dalam mengikuti pembelajaran. Pembelajaran tersebut berfokus kepada pada System & Application Development dan Business Intelligence, seperti Database & SQL, Python for

Data Analytics, ETL Pentaho, Statistic, Data Visualitzation (Tableau & Power BI), Big Data & Cloud Technology, Machine Learning, Deep Learning, Data Warehouse

Program ini dilaksanakan online, secara synchronous dan asynchronous dalam 16 kali pertemuan (1 semester). Setiap modul pembelajaran yang disediakan dibentuk khusus agar dapat memahami konsep dalam big data, serta soft skills dan hard skills apa yang diminati oleh perusahaan-perusahaan, dan mempraktekkan secara langsung melalui pembelajaran coding dan kegiatan proyek kelompok dan individu. Pada akhir pembelajaran, setiap peserta akan mengikuti assesment dan terlibat di dalam sebuah Project sebagai syarat untuk memperoleh sertifikat Completion atas program Studi Independen tersebut.

2.4. Jadwal Kerja

Adapun Rangkuman Jadwal kerja sebagai berikut:

1. Trainer session adalah sesi belajar-mengajar yang dilakukan dalam kelas secara online, pada waktu yang telah disepakati. Trainer session wajib diikuti oleh seluruh peserta. Dilakukan pada 3x seminggu.
2. Mentor Session adalah sesi pembelajaran & tanya jawab yang dilakukan antara pengajar dan peserta secara online pada waktu yang telah disepakati di luar jam belajar mengajar. Mentor session dilakukan hanya beberapa peserta saja.

BAB III

Celerates Acceleration Program (Big Data & Business Intelligence)

3.1. Deskripsi Projek Akhir

Credit risk analysis adalah resiko yang harus ditanggung oleh seorang individu atau lembaga ketika memberikan pinjaman - biasanya dalam bentuk uang - ke individu atau pihaklain. Resiko ini berupa tidak bisa dibayarkannya pokok dan bunga pinjaman, sehingga mengakibatkan kerugian berikut: gangguan aliran kas (cash flow) sehingga modal kerja terganggu. meningkatkan biaya operasional untuk mengejar pembayaran tersebut (collection). Untuk memperkecil resiko kredit ini, biasanya dilakukan proses yang disebut dengan credit scoring dan credit rating terhadap pihak peminjam. Output proses ini akan menjadi basis untuk menentukan apakah aplikasi pengajuan pinjaman baru diterima atau ditolak.

Credit risk adalah bentuk resiko kegagalan bayar oleh debitur yang ditanggung oleh creditor. Resiko ini bisa berdampak pada: Penurunan income dari bunga Gangguan modal kerja yang tidak kembali * Meningkatnya biaya yang berasal dari collection Untuk memperkecil resiko tersebut, sebuah praktik yang wajar di seluruh perusahaan keuangan untuk melakukan credit scoring dan credit rating pada calon debiturnya. Output dari sistem scoring akan menjadi tolak ukur apakah pengajuan pinjaman bisa diproses atau tidak.

Credit score adalah nilai yang diberikan pada calon debitur berdasarkan history dan/atau parameter-parameter tertentu. Proses dari pemberian credit score ini disebut sebagai credit scoring, berarti calon debitur dinilai layak untuk diberi pembiayaan atau pinjaman. (Karena dianggap mampu mengembalikan pinjaman)

Risk rating; kadang banyak lembaga yang juga menggunakan risk rating. Berbanding terbalik dengan credit scoring, semakin tinggi nilai risk rating menunjukkan bahwa debitur semakin berisiko.

3.2. Metode Penelitian

Tugas akhir ini ditujukan untuk membuat alat yang dapat membantu, khususnya area. Untuk mencapai tujuan itu, metode data analisis dan visualisasi yang kreatif dilakukan pada riset ini. Beberapa fase yaitu pengumpulan, pemrosesan, pembersihan, eksplorasi analisis data, dan pemodelan

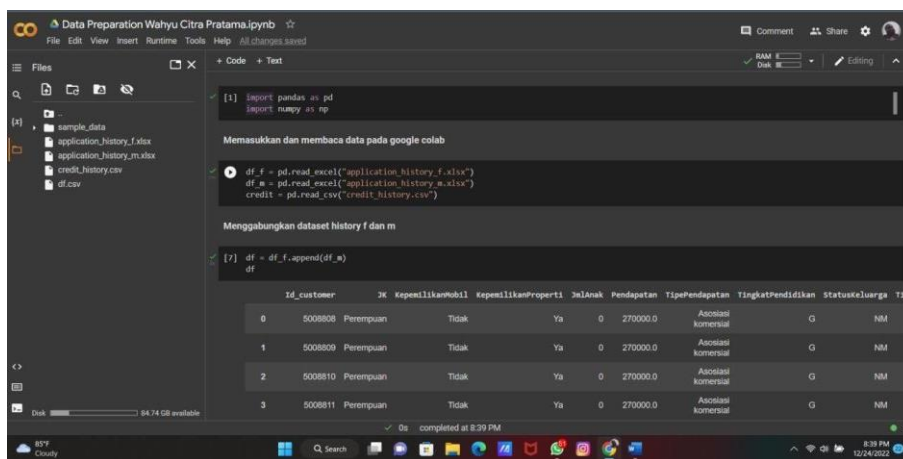
algoritma diperlukan untuk mendapatkan hasil yang diinginkan. Selanjutnya, analisis data digunakan untuk membuat platform peta interaktif.

3.3. Proses Pelaksanaan Projek

Adapun proses dari projek akhir sebagai berikut

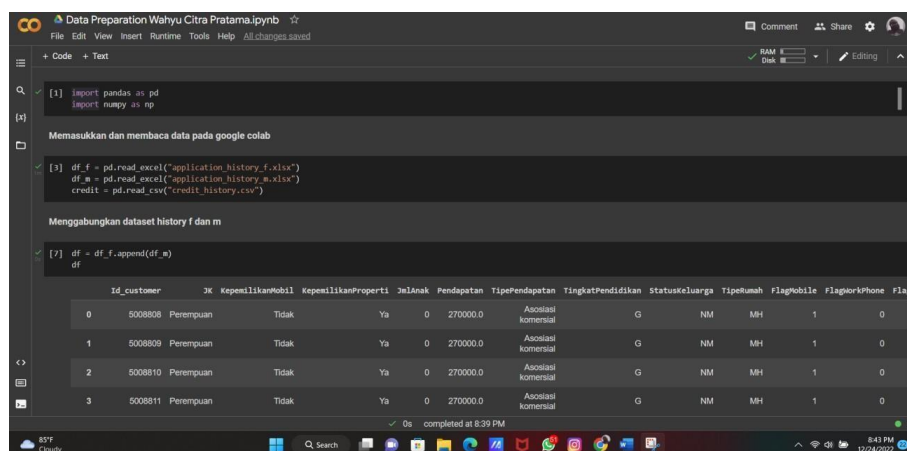
3.3.1. Analisis Data

Pertama dataset yang telah diberikan di soal didownload dan di upload data set pada pengerjaan di Data preparation tersebut yaitu data set Application_history_f.xls, Application_history_m.xls, credit.history.csv.



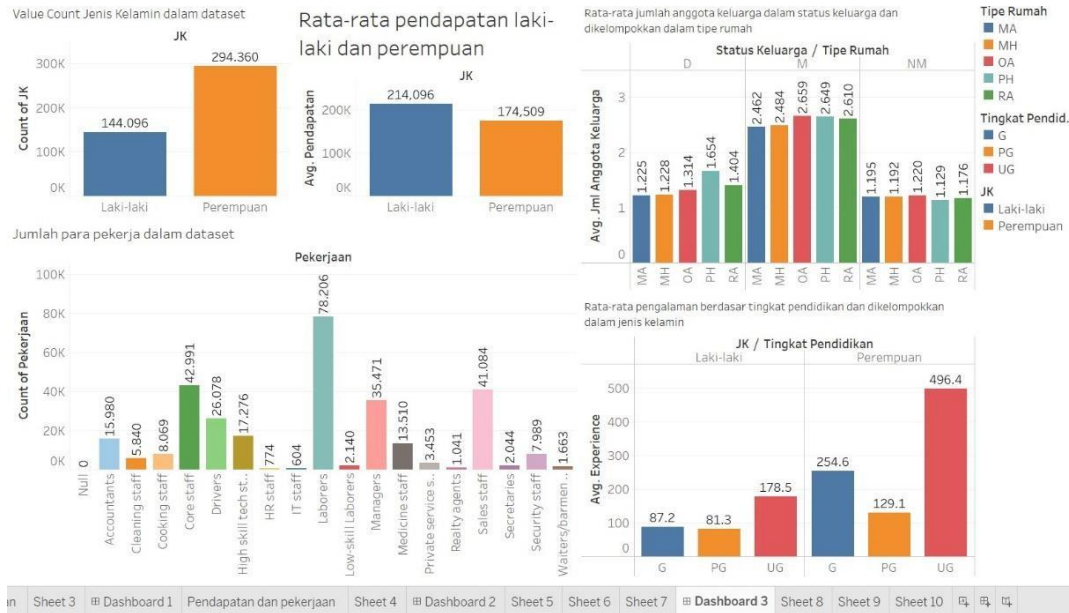
Gambar 1 Analisis Data 1

Link:[https://colab.research.google.com/drive/1XrrGsXJlVI1KZfJP9ulCDvYgY_MFwIWQ?usp=sha](https://colab.research.google.com/drive/1XrrGsXJlVI1KZfJP9ulCDvYgY_MFwIWQ?usp=sharing)
[ring](https://colab.research.google.com/drive/1XrrGsXJlVI1KZfJP9ulCDvYgY_MFwIWQ?usp=sharing)



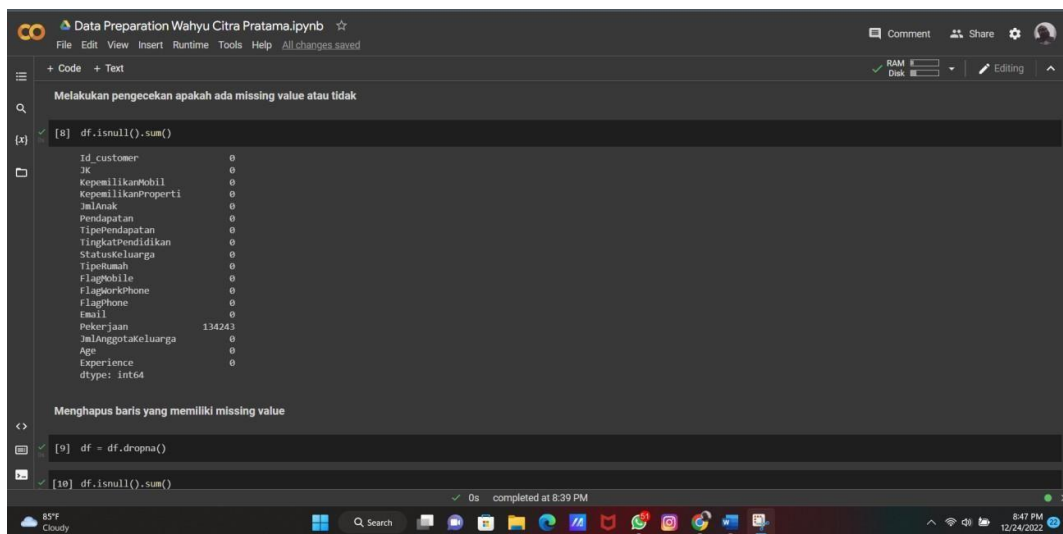
Gambar 2 Analisis Data 2

Dan selanjutnya setelah data tersebut dimasukan data di coding menggunakan aplikasi Google Colab yang dimana saya menggunakan codingan Bahasa python dan selanjut nya data dieksekusi dimana awal codingan kita menggabungkan dataset history female(f) dan male(m). seperti gambar diatas.



Gambar 3 Analisis Data 3

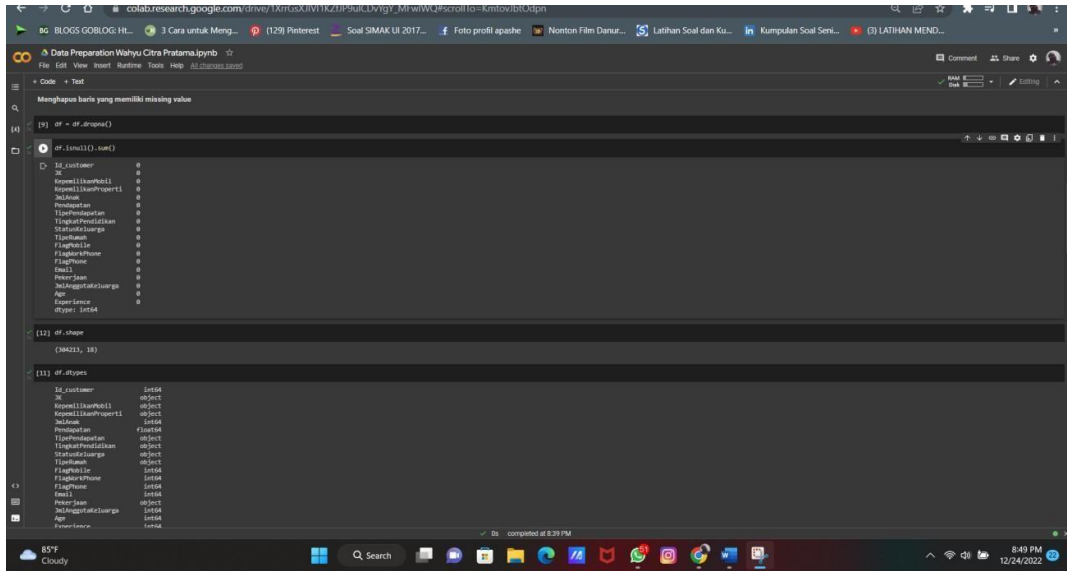
Melakukan analisis data pada tablu menggunakan dataset history_aplikasi yang sudah digabung. Bisa dilihat bahwa terdapat missing value pada kolom pekerjaan. Dan tipe data yang dibagi menjadi 2, yaitu numerical dan string.



Gambar 4 Analisis Data 4

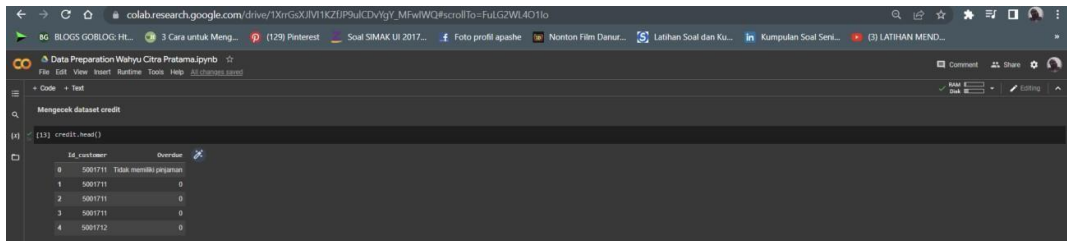
Dan selanjutnya setelah kita menggabungkan dataset f dan m, kita melakukan pengecekan pada

data apakah ada data missing value atau tidak nya.

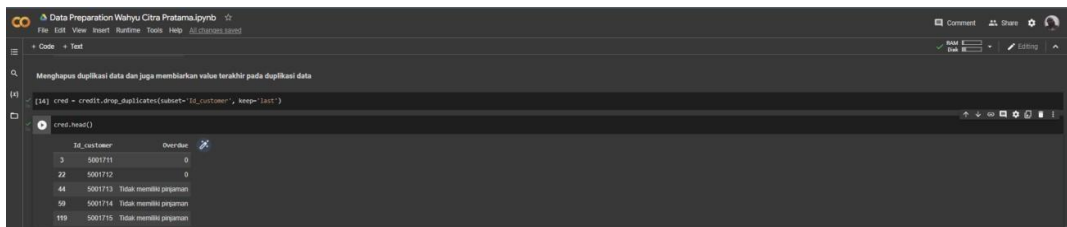


Gambar 5 Analisis Data 5

Selanjutnya data yang telah kita cek data valuenya kita menghapus missing value karena akan mengganggu proses modelling. Dan dibawah nya agar kita melihat jumlah data yang ada pada dataset dan tipe data dari data set df.

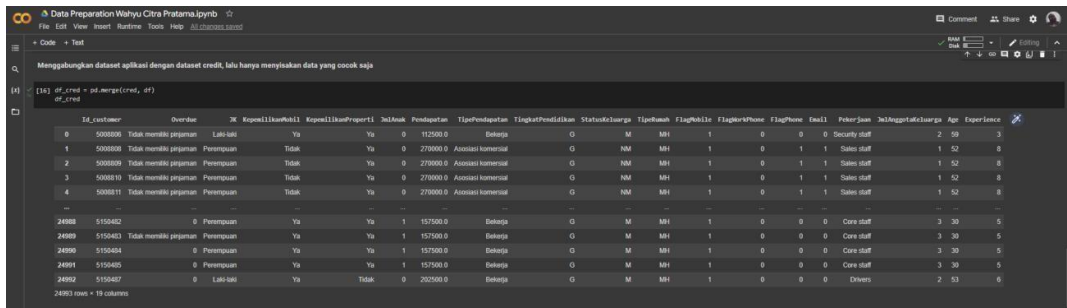


Gambar 6 Analisis Data 6



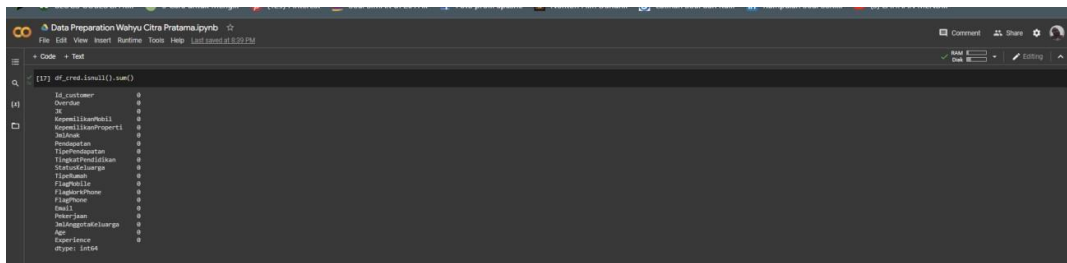
Gambar 7 Analisis Data 7

Melihat dataset credit dengan menampilkan 5 data teratas pada dataset tersebut dan Menghapus duplikasi data pada dataset credit yaitu pada kolom Id_customer tetapi membiarkan value terakhir pada duplikasi data tersebut dengan fungsi drop_duplicates().

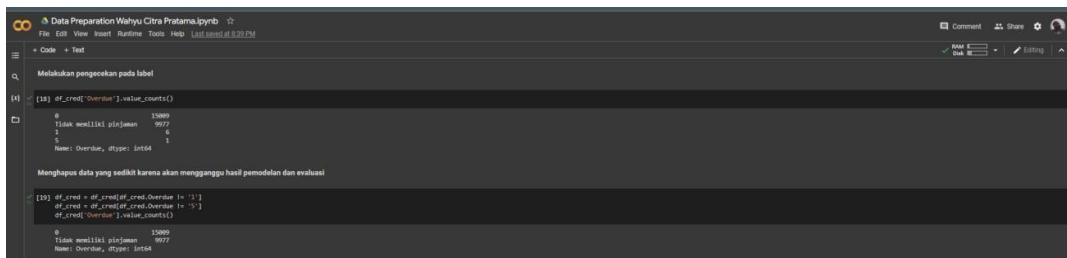


Gambar 8 Analisis Data 8

Menggabungkan dataset history aplikasi dan dataset credit, namun hanya menggabungkan data dengan kecocokan yang sama dan data yang tidak sama akan dihapus. Lalu dataset tersebut akan diberi nama variabel baru.

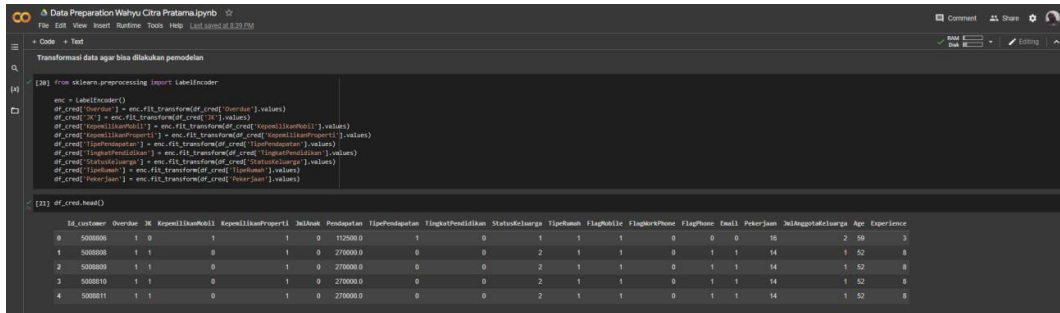


Gambar 9 Analisis Data 9



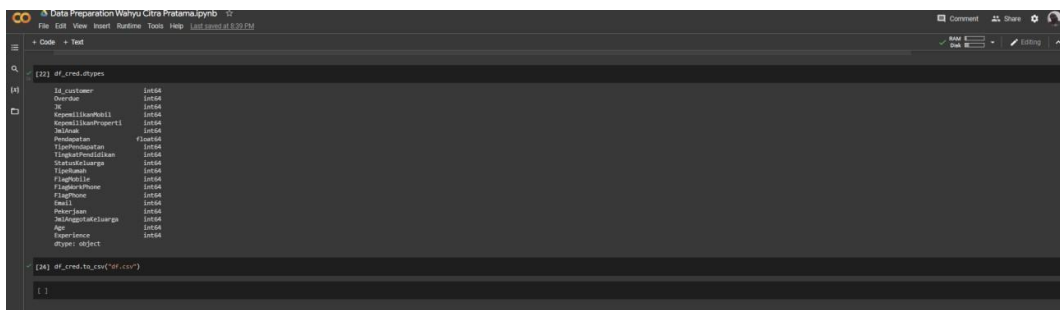
Gambar 10 Analisis Data 10

Melakukan pengecekan missing value dan Melakukan pengecekan label yaitu pada kolom overdue dengan mengecek jumlah value, lalu menghapus data dengan jumlah value yang sedikit, karena akan mengganggu proses pemodelan dan evaluasi pemodelan. Setelah itu dataset akan di save untuk dilakukan visualisasi pada tableu.



Gambar 11 Analisis Data 11

Transformasi data menggunakan fungsi yang ada pada library sklearn. preprocessing yaitu labelencode. Data ini akan diubah menjadi bentuk data numerical, agar dataset tersebut bisa digunakan dalam pemodelan.

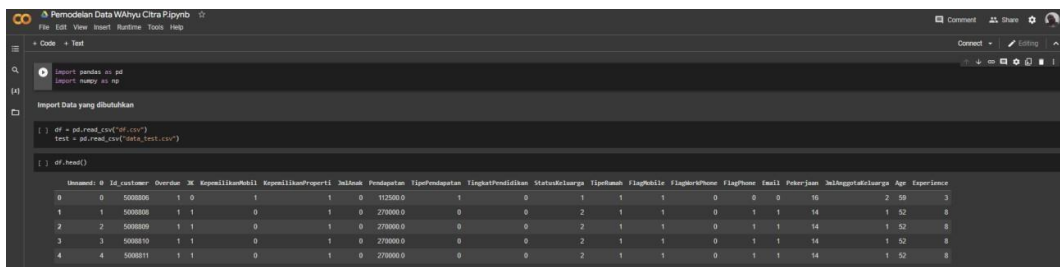


Gambar 12 Analisis Data 12

Pengecekan tipe data df_cred dan didapatkan data sudah berubah semua ke dalam bentuk numerical. Lalu menyimpan dataset untuk digunakan dalam proses pemodelan.

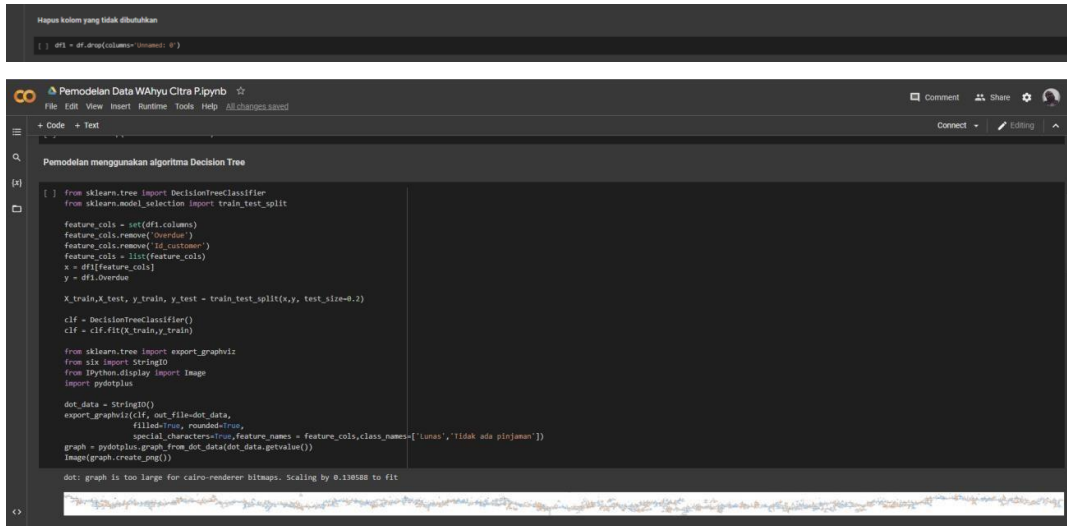
3.3.2. Pengolahan Data

Link: <https://colab.research.google.com/drive/159bjwOk8LhZ1KiHPnCARS49b45Gkbt9-?usp=sharing>



Gambar 13 Pengolahan Data 1

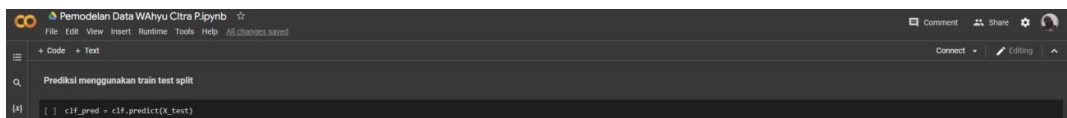
Import library dan juga import data ke dalam colab, kemudian melakukan pengecekan data dengan melihat 5 data teratas pada dataset df dan Menghapus Kolom yang tidak dibutuhkan.



Gambar 14 Pengolahan Data 2

Selanjutnya pemodelan

- 1) Pemodelan dimulai dengan import library yang dibutuhkan dari sklearn yaitu DecisionTreeClassifier untuk algoritma pemodelan dan juga train_test_split untuk melakukan split data.
- 2) Mendefinisikan variabel x dan y, yaitu x dengan dataset tanpa kolom overdue dan id_customer, sementara y dengan kolom overdue yang ada pada dataset saja.
- 3) Lalu melakukan split data antara data train dan data test dengan rasio 8:2
- 4) Fit data train ke dalam model data yang sebelumnya sudah didefinisikan dalam variabel clf
- 5) Lalu terakhir pemodelan data dan menampilkan pemodelan tersebut, yaitu hasil pemodelan dengan algoritma decision tree



Gambar 15 Pengolahan Data 3

Melakukan prediksi menggunakan algoritma decision tree dengan data test X


```

[] cols = ['id_customer', 'Value']
Submission = test[cols]
Submission.head()

id_customer  Value
0  5142248      1
1  5036925      0
2  5126080      1
3  5088887      1
4  5022156      0

[] Submission.columns = Submission.columns.str.replace('id_customer', 'Id')
Submission.head()

Id  Value
0  5142248      1
1  5036925      0
2  5126080      1
3  5088887      1
4  5022156      0

Submission

[] Submission.to_csv("Submission.csv")

```

Gambar 19 Pengolahan Data 7

Memasukkan kolom id_customer dan value ke dalam dataset submission untuk kebutuhan tugas

```

[] from sklearn.metrics import f1_score, accuracy_score, classification_report, confusion_matrix

print("F1 Score : ", f1_score(y_test, clf_pred))
print("Accuracy Score : ", accuracy_score(y_test, clf_pred))
print("Classification Report :\n", classification_report(y_test, clf_pred))
print("Confusion Matrix :\n", confusion_matrix(y_test, clf_pred))

F1 Score : 0.6182035228848822
Accuracy Score : 0.718887558220809
Classification Report :
      precision    recall  f1-score   support

 0       0.75     0.81     0.78     3859
 1       0.66     0.57     0.61     1939

 accuracy: 0.78
 macro avg: 0.69
 weighted avg: 0.71     0.72     0.71     4998

Confusion Matrix :
[[2483  560]
 [ 630 1180]]

```

Gambar 20 Pengolahan Data 8

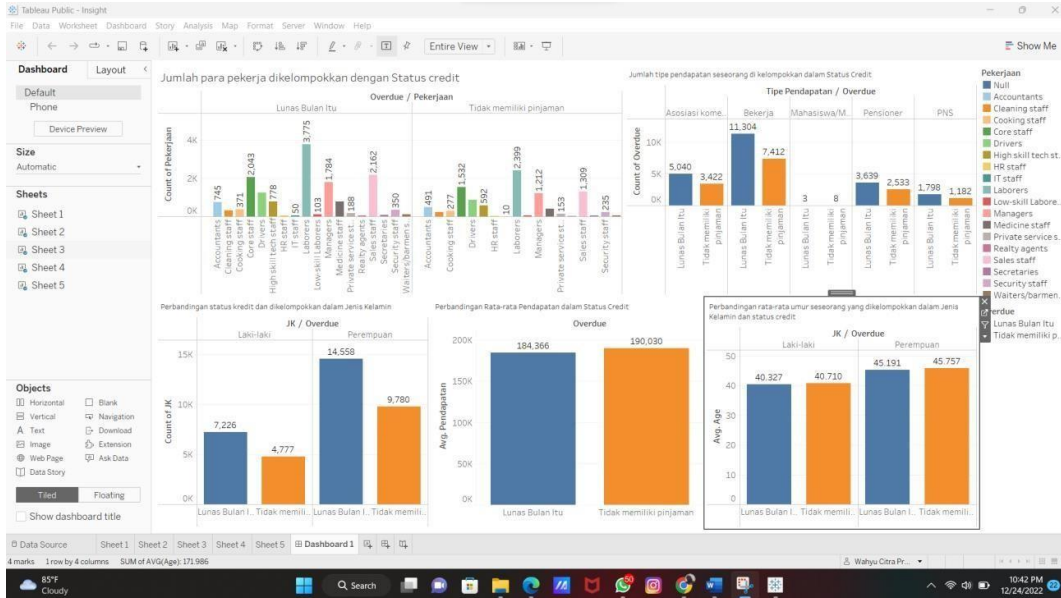
Evaluasi hasil prediksi yang sudah dilakukan sebelumnya menggunakan F1 Score, Accuracy Score, Classification Report dan Confusion Matrix.

3.4. Hasil Pencapaian Proyek

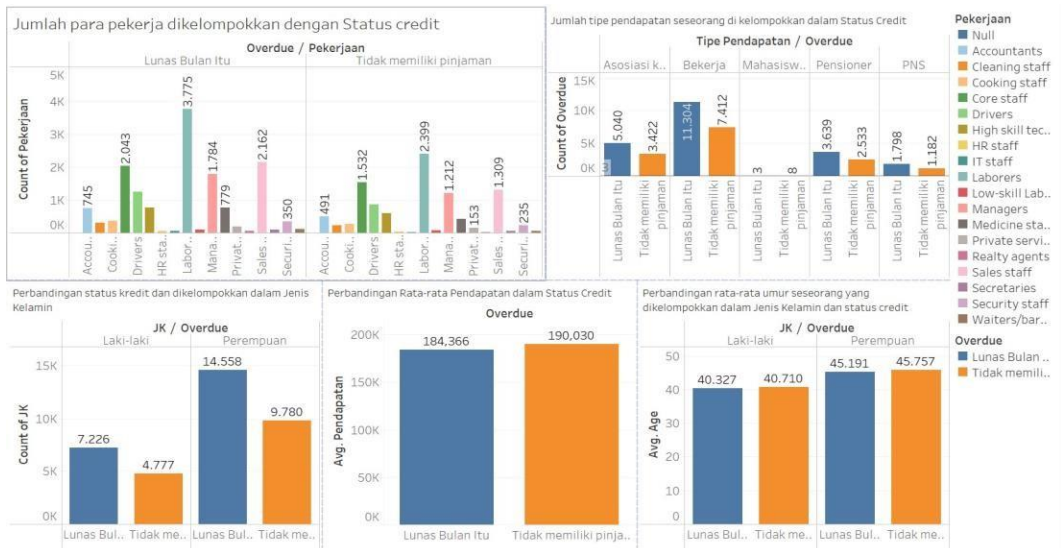
Berikut adalah hasil pencapaian dari proyek akhir

3.4.1. Visualisasi Data

Link: https://public.tableau.com/views/Insight_16718964946880/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link



Gambar 21 Visualisasi Data 1



Gambar 22 Visualisasi Data 2

Melakukan visualisasi data menggunakan tableau untuk dilihat informasi yang berguna, dimana pada sheet pertama, didapatkan informasi bahwa pekerjaan labores merupakan pekerjaan paling tinggi yang memiliki pinjaman namun lunas pada bulan itu, begitu juga pada bagian tidak memilikipinjaman. Lalu selanjutnya pada kolom disebelah kanan atas, terdapat informasi bahwa seseorang dengan tipe pendapatan bekerja memiliki nilai yang tinggi dalam status pinjaman, baik itu lunas ataupun tidak ada pinjaman. Kemudian perempuan banyak yg melakukan pinjaman namun lunas pada bulan itu. Ada juga rata-rata pendapatan yang melakukan pinjaman lebih rendah daripada yang tidak melakukan

pinjaman. Sementara rata-rata umur seseorang yang melakukan pinjaman lebih rendah dibandingkan dengan yang tidak melakukan pinjaman baik itu perempuan atau laki laki.

Penjelasan Hasil Pengolahan dan Visualisasi Data Link: <https://youtu.be/AwMtrfvzLvk>

BAB IV

Penutup

4.1. Kesimpulan

Kesimpulan Dari penelitian ini, peneliti membuktikan bahwa pengembangan platform visualisasi data dalam pemetaan kreatif adalah metode yang optimal untuk memahami distribusi, manajemen, dan mengoptimalkan sistem pada analysis risk tersebut. Melalui data yang tersedia dan pengerjaan dalam tugas akhir juga memungkinkan pada pengguna antara laki laki dan perempuan.

4.2. Saran

Saran Adapun saran mengenai proses pelaksanaan pengerjaan tugas akhir ini yaitu:

- 1) Gunakan Alias atau Kode untuk Variabel Data Jika data tidak memiliki struktur pada langkah pertama, tidak ada sistem yang memberitahu data apa yang harus dikumpulkan. Masalah ini dapat mengarah ke duplikasi dan pengulangan variabel dalam data. Hal ini dapat terjadi dalam satu dataset dan itu akan menjadi jauh lebih sulit dalam kombinasi dataset yang berbeda. Lebih baik untuk memiliki alias atau kode untuk variabel tertentu untuk mencegah beberapa kesalahan variabel. Hal ini juga memudahkan untuk menghitung atau mengelola variabel data dengan solusi ini.
- 2) Meningkatkan Proses Pengumpulan Data Proses pengumpulan data perlu ditingkatkan untuk mencegah banyak data yang tidak lengkap sehingga dapat menyebabkan kesalahan visualisasi data. Nilai kosong di dalam data benar-benar memengaruhi proses analisis data. Ini perlu dicegah saat data dikumpulkan dengan memberikan batasan agar data tidak diperbolehkan dikirimkan jika belum lengkap

BAB V

Lampiran A. TOR



• **PT Mitra Talenta Grup**
The Manhattan Square Building, Mid Tower, 12th Floor.
Jl. T.B. Simatupang kav 1-5, Cilendak Timur, Jakarta Selatan

• **Celerates.com**
contact@celerates.com
021 8064 1000

TERM OF REFERENCE (TOR)

STUDI INDEPENDEN CELERATES ACCELERATION PROGRAM (BIG DATA & BUSINESS INTELLIGENCE) PROGRAM MSIB KAMPUS MERDEKA BATCH-3

1. Latar Belakang

Pesatnya perkembangan teknologi, menyebabkan cepatnya perubahan terhadap pembaharuan teknologi.

Sebagai perusahaan yang terlibat di dalam industri teknologi, PT Mitra Talenta Grup (Celerates) ingin berpartisipasi untuk menghasilkan tenaga kerja dengan melibatkan mahasiswa yang berpotensi untuk menjadi tenaga profesional di masa depan yang turut mendukung kemajuan teknologi di Indonesia. Program studi independen menjadi salah satu cara dari PT Mitra Talenta Grup (Celerates) untuk mencapai misi tersebut.

2. Tujuan Kegiatan

- a. Mahasiswa dibekali oleh ilmu-ilmu yang diberikan oleh mentor yaitu praktisi di dunia IT.
- b. Mahasiswa dibekali ilmu melalui praktek dari teori yang sudah dibekali sebelumnya.
- c. Mahasiswa dibekali soft skill untuk menghadapi dunia kerja setelah lulus dari program MSIB.
- d. Mahasiswa memiliki hasil dari Project Akhir untuk digunakan sebagai bekal portofolio.

3. Periode & Tempat Pelaksanaan

- a. Periode : 18 Agustus 2022 - 31 Desember 2022
- b. Tempat : Online melalui Zoom

4. Peserta

Jumlah peserta kegiatan ialah 88 orang, yaitu mahasiswa yang terpilih untuk mengikuti program kami yang tersebar dari seluruh Indonesia.

5. Penyelenggara

Penyelenggara kegiatan adalah PT Mitra Talenta Grup (Celerates).

6. Penutup

Demikian Term of Reference kegiatan Studi Independen Celerates Acceleration Program (Big Data & Business Intelligence). Untuk informasi lebih lanjut, silahkan menghubungi penyelenggara kegiatan.

BAB VI

Lampiran B. Log Activity

Minggu/Tgl	Kegiatan	Hasil
Minggu ke-1 18 - 19 Agu 2022	Live session	<p>Dimulai dengan pembelajaran syncronus dengan trainer agar salaing mengetahui antara trainer dan mitra dimana mempelajari tentang pengenalan awal SQL yaitu mempelajari perbedaaan pada database seperti query mencari number nanti yang akan dipake untuk materi dan project selanjutnya.</p> <p>Pembelajaran pada sesi awal pembelajaran kali ini kami disajikan quiz dan ppt pemebelajaran untuk mengetahui materi yang kemarin dijabar kan dan selanjutnya mempelajari tentang materi pemebelajaran awal SQL mendownload dan menginstall aplikasi yang bernama DBeaver dimana pada sesi ini kita belajar langsung kedatabase nya dan mengesekusi pada databasenya pada database nortwind degan host yang sduah diberikan. dan memberikan penjelasan pada tabel pada database nortwind tersebut</p>
Minggu ke-2 22 - 26 Agu 2022	Live session dan mentoring	<p>Pada hari Senin, 22-08-2022 Mentor: Bpk. Asfar pada sesi kali ini kami di beri materi untuk mengisi pada web miro untuk mengisi pertanyaan seputar data itu apa? dan apa yang nanti nya di dapat dari data? dan sebagainya.</p> <p>Pada hari Selasa, 23-08-2022 Mentor: Bpk Afzar Jaya .pembelajaran pada sesi awal pembelajaran kali ini kami disajikan materi serta quiz yang perlu dikerjakan seberapa mana kita mengerti mengenai database tersebut. Apa yang kamu kerjakan hari ini?</p> <p>Pada hari rabu, 24-08-2022 Mentor: Bpk Asfar Jaya . pada sesi pembelajaran kali ini kami disajikan materi serta aplikasi postgresql yang perlu diinstall dan dipelajari serta kita memahami dari aplikasi tersebut.</p> <p>Pada hari kamis, 25-08-2022 pada sesi kali ini kami dimulai dari mengisi pretest dan kuis, kemudian dilanjutkan dijelaskan nya - Apa itu ERD - Kegunaan ERD - Komponen dan Konsep dasar yang terdapat pada ERD ; Relasi,atribut,entetitas,strong and weak entetitas - dan yang</p>

		<p>terakhir cara membuat ERD itu sendiri (mengidentifikasi entitas, hubungan, kardinalitas dan indentifikasi atribut.</p> <p>Pada hari Jumaat, 26-08-2022 pada sesi kali ini kami dimulai dari mengisi pretest dan kuis, kemudian dilanjutkan - dijelaskan nya pelajaran yang kemarin dan membuat table dari data base - Komponen dan Konsep dasar yang terdapat pada data bas</p>
<p>Minggu ke-3 29 Agu - 02 Sep 2022</p>	<p>Live session, Cycle meeting dan Mentoring</p>	<p>Pada hari Senin, 29 Agustus 2022 sesi ini kami mengisi usacase tugas trainer dengan dilist di kelompok masing masing</p> <p>Pada hari Selasa, 30 Agustus 2022 sesi nidiawali dengan pembagian materi dan webinar dari celerates yang mempelajari tentang sebuah bisnis dan data yang dimiliki harus seimbang dengan bagaimana cara menganalisa data secara tepat, otomatis dan kompleks.</p> <p>Pada Hari Rabu, 31 Agustus 2022 kami diawali dengan mempelajari kembali dan mendiskusikan tugas dari trainer dan mempelajari tentang phyton agar bisa mempelajari buat hari selanjutnya yaitu pada hari kamis tentang phyton itu sendiri.</p> <p>Pada hari Kamis, 1 September 2022 Pada sesi kali ini kami disajikan pembelajaran phton training dimana kita mempelajari bagaimana menginstall dan mempelajari function dari phyton itu sendiri dan dimana selnjutnya diakhiri dengan kuis yang disajikan dari trainer itu sendiri.</p> <p>Pada hari Jumaat, 2 September 2022 pada sesi ini kami Pada hari jumaat tanggal 2 september 2022, pada sesi kali ini kami melanjutkan dan mempelajari phyton training dan kami di sajikan exercise seperti soal dan dibuat kelompok untuk mengerjakan soal tersebut yng diberikan oleh trainer dan dikumpulkan</p>
<p>Minggu ke-4 05 - 09 Sep 2022</p>	<p>Live session, Talk show, Cycle meeting dan Mentoring</p>	<p>Pada hari Senin, 5 September 2022 Pada sesi kali ini kami di beri materi lanjutan tentang phyton training mapel ini menjelaskan dan melanjutkan apa itu tipe data di statistik, sampling ,balance dan imbalanced data dan diakhiri dengan membuat kolompok dimana menjawab soal dari trainer.</p> <p>Pada hari Selasa, 6 September 2022 pada sesi ini kami memahami kembali materi sql selanjutnya mentor memberikan kami tugas yang harus dikerjakan yaitu dengan tugas study case yang di dalam nya berisi data data.</p>

		<p>Pada hari Rabu, 7 September 2022 , Pada sesi ini kami diberikan dan diajarkan kembali tentang phyton for analythics yang didalam nya mengajarkan tentang machine learning, matrix performance , data statitistik</p> <p>Pada hari Kamis, 8 September 2022 pada sesi kali ini kami diadakan sesi mentoring dimana kita di beri kesempatan untuk memberikan pertanyaan tentang materi kemarin yang dijaarkan yaitu tentang statistika dan phtyon analisis</p> <p>Pada hari Jumaat, 9 September 2022 kami diajarkan awal pengenalan statistik dimana daa nya explorasi data analisis, metode sampling, histogram dan lain lain</p>
Minggu ke-5 12 - 16 Sep 2022	Live session, Learning motivated, Cycle meeting dan Mentoring.	<p>Pada hari Senin, 12 September 2022, kami mempelajari kembali Statistik data dimana didalam nya ada Exploratory Data analisis sebuah proses kritis dalam melakukan investigasi awal pada data dengan tujuan menemukan pola, anomali, menguji hipotesis dan dapat memeriksa asumsi dengan bantuan statistik ringkasan kemudian representasi grafis (visualisasi). Sampling adalah bagian dari metodologi statistika yang berhubungan dengan prosedur sistematis dalam pengambilan sebagian anggota populasi untuk keperluan pendugaan. Histogram alam ilmu statistika, pengertian histogram adalah representasi grafis (diagram) yang mengatur dan menampilkan frekuensi data sampel pada rentang tertentu. Frekuensi data yang ada pada masing-masing kelas direpresentasikan dengan bentuk grafik diagram batang atau kolom. Prinsip kerja histogram ialah dengan menjadikan sekumpulan data sampel penelitian menjadi bentuk visual yang mudah ditafsirkan dengan mengambil titik data kemudian mengelompokkannya kedalam rentang yang logis.</p> <p>Pada hari Selasa,13 Septe,ber 2022 .Kali ini kami mengikuti mentoring class dimana kita disuruh membahas tentang pelajaran yang kita ikutin kemarin apakah sudah paham atau tidak nya kami dan kami di kasih pertanyaan dari mentor untuk kita mengetahui sejauh mana dan paham tentang pelajaran yang kami ikuti.</p> <p>Pada hari Rabu,14 September 2022, dimana pada sesi ini kami mempelajari tentang "Big Data Sience" dan keykonsep nya. Selanjut dimana kita mengikuti Praktek nya dari proses data since nya , past to future, business fact base nya dan mempelajari Busines Statistik yang mempelajari cara perhitungan tersebut. Pada hari Kamis,15 September 2022, kami mempelajari kembali big data since dimanna kami</p>

		<p>mempelajari poulasi kumpulan dari sekelompok objek,individu. atau entitas serupa yang dapat diamati didalam nya bagian dari populasi yang ditarik menurut aturan atau rencana untuk menyimpulkan karakteristik disebut sampel.</p> <p>Pada hari Jumaat.16 September 2022, kami mempelajari Data- Information-Knowledge-Wisdom atau yang disebut dengan Dikw Pyramid data adalah fakta dan angka mentah yang dengan sendirinya tidak memiliki arti data perlu diubah menjadi informasi yang berarti dan disajikan dalam format yang paling berguna dan harus diproses dalam konteks untuk memberikan makna . information dalah dan dan makna. Knowledge(pengetahuan) adalah langkah selanjutnya dan paling signifikan. dapat mengambil data, mengategorikan dan mengolahnya menghasilkan informasi. Kemudian mengatur semua informasi kemudian mengatur semua informasi ini sedemikian rupasehingga dapat bermanfaat.. sedangkan informasi membantu kita untuk memahami hubungan dan konteks . Wisdom adalah batas akhir memprediks dengan benar, tidak hanya dengan mendekteksi dan memahami pola dan mempelajari juga secara dalam.</p>
<p>Minggu ke-6 19 - 23 Sep 2022</p>	<p>Live session,Talksho w, Learning motivated, Cycle meeting dan Mentoring.</p>	<p>Pada hari Senin, September 2022 sesi kali ini mempelajari tentang: - Overview oltp vs OLAP - Introducing Data Warehouse Concept - ETL in Data Warehouse ETL didefinisikan sebagai proses yang mengekstrak data dari sistem sumber RDBMS yang berbeda, kemudian mengubah data (seperti menerapkan perhitungan, penggabungan, dll.) dan akhirnya memuat data menjadi sistem Data Warehouse. Ada dua jenis metode ekstraksi data warehouse: • Metode Ekstraksi Logis • Metode Ekstraksi Fisik Data Warehouse adalah sebuah sistem yang bertugas mengarsipkan sekaligus melakukan analisis data historis untuk menunjang keperluan informasi pada sebuah bisnis ataupun organisasi.</p> <p>Pada hari selasa 20 September 2022, pada sesi ini saya tidak mengikuti mentoring dan kami mempelajari yang diberikan mentor yaitu mempelajari kembali pelajaran yang diiikuti kemarin dn sesi tanya jawab</p> <p>Pada Hari Rabu, 21 September 2022, sesi kali ini kami diberi materi tentang ETL vs ELT dimana adanya perbedaan antara etl dan elt ETL : - Data ditransformasikan di server pementasan dan kemudian ditransfer ke gudang data DB. -</p>

		<p>Transformasi dilakukan di server ETL/area pementasan. dan selanjutnya kami mempelajari bagian dalam etl tersebut dan tidak lupa dengan ELT tersebut di aplikasi pentaho. - Data pertama dimuat ke staging dan kemudian dimuat ke target sistem. Waktu yang intensif. - Proses ETL perlu menunggu transformasi selesai. Sebagai ukuran data bertambah, waktu transformasi bertambah. - Perlu perawatan tinggi karena Anda perlu memilih data untuk dimuat dan mengubah. ELT: - Data tetap berada di DB Data warehouse. - Transformasi dilakukan di sistem target - Data dimuat ke sistem target hanya sekali. Lebih cepat. -Dalam proses ELT, kecepatan tidak pernah bergantung pada ukuran dari datanya. - Perawatan rendah karena data selalu tersedia.</p> <p>Pada Hari Kamis. 22 September 2022, pada sesi kali ini kami mempelajari dan training pentaho apa yang diberikan materi sebelumnya dan adanya sesi tanya jawab tentang sejauh mana kami mempelajari tentang training tersebut dan aplikasi pentahonya.</p> <p>Pada Hari Jumaat, 23 September 2022, pada sesi kali ini kami mempelajari tentang : - Pentaho Data Integration Installation - Intro Pentaho Data Interegration - Pentaho Data Intregation Components - PDI input dan Output dan selanjutnya dimana kami mmempelajari tentang pentaho dari pengenalan data grid dan sampai pengenalan dari input dan ouput dari aplikasi pentaho tersebut.</p>
Minggu ke-7 26 - 30 Sep 2022	Live session, Cycle meeting, dan Mentoring	<p>Pada Hari Senin, 26 September 2022. pada sesi ini kami diberikan materi untuk dipelajari yaitu materi tentang ; - Testing ETL using generate rows data - Load Multi Files to Database - add Constans Sequerces - replace String Value dimana sesi ini kami mempelajari dimana penggunaan rows data pada etl dan dilanjutkan dengan menggunakan aplikasinya untuk kami coba dan selanjutnya load multi files adalah diman tempat uoload dile sederhana tanpa menyimpan di database.</p> <p>Selasa 27 Sep 2022 Apa yang kamu kerjakan hari ini? Pada hari Selasa, Tanggal 27 September 2022. pada sesi mentoring kali ini kami mempelajari tentang activity flow dan mempelajari kembali pelajaran dari - Load Multi Files to Database - add Constans Sequerces - replace String Value sembunyan Rabu 28 Sep 2022</p> <p>Pada Hari Rabu, 28 September 2022, pada sesi ini mempelajari tentang flow dari pentaho tersebut selanjutnya</p>

		<p>mengerjakan yang diberi file oleh trainer dengan data input output nya dari data grid. dan memasukan data seperti average <60 : bad average >= 60 dan <= 70 : Average Average >= 70 dan <= : Good Average > 80 : Exeellent</p> <p>Pada hari Kamis ,29 September 2022, pada sesi kali ini kami mempelajari Asyhronus dan disertai apa saja yang pelajari dibuu introduction businees Analystics Using Simulation- dan memberikan landasan bagi pengambilan keputusan ekonomi manajerial dan untuk selanjutnya topik seperti manajemen risiko dan opsi nyata . berikut tiga perangkat dasar analisis bisnis yang disajikan dalam buku ini . 1. penataan masalah keputusan 2. mengukur ketidak pastian Analisis Deskriptif analisis probalitas dan statistik 3. menentukan keputusan yang optimal Preciptive Analystics simulasi dan optimasi analisis bisnis akan dibagi menjadi tiga topik beas descriptivenya, analytics dll.</p> <p>Pada hari Jumaat,30 September 2022, pada sesi kali ini kami melakukan Etl dengan data grips, add constants, calculator range dan table output</p>
<p>Minggu ke-8 03 - 07 Okt 2022</p>	<p>Live session, Cycle meeting, dan Mentoring.</p>	<p>Pada hari Senin, 3 Oktober 2022. Pada sesi kali ini kami mempelajari dengan mempraktekan Bloks this stel until finish, Filter Rows,Switch/ Case dan join di pentaho. dan dijelaskan juga bagaimana real case proses ELT. dan kami juga mempelajari apa itu Join pada proses tersebut dengan dijelaskan Tipe - Tipe dari Join - Inner Join - Outer Join Left Outer Right Outer Full outer - CArtesian Join Pada hari</p> <p>Selasa, 4 Oktober 2022. pada sesi ini melakukan Merge row diff di pentaho. sebelum itu diketahui apa saja terdapat pada merge rows(diff) flags indetical changed, new dan delete</p> <p>Pada hari Rabu, 5 Oktober 2022. Pada sesi kali ini mai mempelajari pembelajaran sinkron dengan relatime yaotu pembelajaran melanjutkan merge rows (diff) flags di pentaho tersebut yang dilakukan dengan flags identical,changged,new dan delete pad tada dan table di pentaho</p> <p>Pada hari Kamis,6 Oktober 2022. pada sesi kali ini kami mempeladi pembelharan dengan mentor yang menjelaskan kembali pelajaran dari trainer kemarin dan mempraktekan kembali yang dijelaskan dengan stream lookup orders stream lookup, shipment dan dilanjutkan juga dengan sesi jawab yg dimana kami mengikuti pelajaran penaho tersebut berserta database nya.</p>

		<p>Pada hari Jumaat, 7 Oktober 2022. Pada sesi kali ini kami melakukan database join, data validator, dan mengupdate table meski terjadi kendala teknis antara trainer dan mentor dan dilanjutkan juga transformation melalui transformasi baru yaitu (mergediff), menambahkan job ke dalam canvas pentaho tersebut.</p>
<p>Minggu ke-9 10 - 14 Okt 2022</p>	<p>Briefing role, Live session, Mentoring, Cycle meeting.</p>	<p>Pada hari Senin, 10 Oktober 2022, pada sesi kali ini kami diberikan absen serta pertanyaan tentang pelajaran yang akan dibahas besok yaitu pembelajaran pada aplikasi Tableau.</p> <p>pada sesi kali ini kami diberikan absen serta pertanyaan tentang pelajaran yang akan dibahas besok yaitu pembelajaran pada aplikasi Tableau.</p> <p>Pada hari Rabu, 12 Oktober 2022. pada sesi ini kami mempelajari tentang tableau tampilan awal dan penggunaannya pada aplikasi tersebut dan diberikan nya tugas untuk mengerjakan secara berkelompok</p> <p>Pada sesi kali ini kami mempelajari kembali tentang penggunaan tableau diberikan nya sample store dimana kita dipelajari</p> <p>Pada hari Jumaat, 14 Oktober 2022. Pada sesi kali ini kami mempelajari kembali tentang penggunaan tableau diberikan nya sample store dimana kita dipelajari dari sample, show me dan lain nya</p>
<p>Minggu ke-10 17 - 21 Okt 2022</p>	<p>Pengerjaan Final project, Cycle meeting, dan Mentoring.</p>	<p>Pada Hari Senin, 17 Oktober 2022. Pada sesi kali ini kami diberikan tugas berkelompok untuk mengerjakan hasil yang kemarin dipelajari oleh trainer dan mempresentasikan tugas tersebut.</p> <p>Pada sesi kali ini kami mempelajari kembali hasil pembahasan dan materi yang di berikan trainer</p> <p>Pada Hari Rabu, 19 Oktober 2022, Pada sesi ini kami dijelaskan tentang grouping pada aplikasi tableau dan dimana disana dijelaskan bagaimana data bisa digrouping dan bisa diubah juga dan gimana kita membuat juga masuk ke database tersebut dengan sample store dari data tersebut</p> <p>Pada hari Kamis, 20 Oktober 2022. pada sesi ini kami bersama mentor diberikan penjelasan tentang data site kemarin dan diberikan tugas kelompok yang di kumpulkan batas waktu 2 November 2022</p>

		Pada hari Jumaat, 21 Oktober 2022. pada sesi ini kami diberikan penjelasan pada tableau dan selanjutnya juga diberikan tugas tableau pada masing masing siswa
Minggu ke-11 24 - 28 Okt 2022	Live session, Cycle meeting, dan Mentoring.	Pada Hari senin, 24 Oktober 2022. Pada sesi ini diawali dengan introduction pada materi data studio dan dimana big query nya juga bisa di gunakan di data studio tersebut. Pembelajaran dilakukan mandiri. Pembahasan yang ringan mengenai The Pyramid Principle. Pentingnya berpikir terstruktur dan komunikasi. Prinsip piramida tidak hanya berguna untuk berkomunikasi dengan eksekutif, tetapi juga efektif untuk berkomunikasi dengan siapa pun yang ingin kita bujuk dengan argumen Memperkenalkan data studio lebih lanjut. fitur yang di dalam nya ada report list. selanjutnya membuat simple reporting dengan data source spreadsheets online_sales (membreak down produk tersebut) bisa mengkostum layout, navigation, dll. membuat scorecard (big numbers) mengubah penampilan dan mengatur style metrics, memperhatikan proporsi, menyesuaikan data series hari ini dilakukan dengan self-regulated learning dan apa saja yang saya pelajari : conditional expressions dan operator (case---> pernyataan ini hanya mengeksekusi kode SQL ketika kondisi tertentu terpenuhi) (null-if--->mengambil 2 input dan mengembalikan NULL jika keduanya sama, jika tidak maka akan mengembalikan argumen pertama yang diteruskan) joins (memahami berbagai jenis join : inner, outer, full dan unions joins pembelajaran dilakukan dengan live session dengan trainer. hal yang dipelajari menambah formula uppercase by category, melakukan formatting style (bisa membuat aturan yang sesuai dengan nya) dan membuat report (data penjualan online) dengan ketentuan menemukan insight dan min 3 chart
Minggu ke-12 31 Okt - 04 Nov 2022	Pengerjaan Final project, Cycle meeting, dan Mentoring.	Pada pertemuan seminggu ini kami membuat tugas dari trainer yang nantinya dikumpulkan di data trainer sebagai nilai. pada hari senin, mempersiapkan data preparation transformation mulai dri csv file input, stream lookup, split fields, select values dan text file output Selasa 1 Nov 2022 mengubah tipe data yang berbeda menjadi tipe data "string" dari csv input carrier, airports, airport dan flight performance_10 menganbil data description dari (csv carrier mengubah nama menjadi Carrier_Name), (csv origin mengubah nama menjadi OriginAirport), (csv ddestination mengubah nama menjadi DestAirport) Rabu 2 Nov 2022 menambah in stream field di replace : to, split data Origin (ORIGINCITY, ORIGINSTATE,

		<p>ORIGINAIRPORT) split data Destination (DESTCITY, DESTSTATE, DESTAIRPORT) Kamis 3 Nov 2022</p> <pre>menambahkan script java script //Script here var DepDelayMinutes; var ArrDelayMinutes; var AirTime; var Flights; var Distance; if (DepDelayMinutes==null) DepDelayMinutes = 0; else DepDelayMinutes=DepDelayMinutes; if (ArrDelayMinutes==null) ArrDelayMinutes= 0; else ArrDelayMinutes=ArrDelayMinutes; if (AirTime==null) AirTime= 0; else AirTime=AirTime; if (Flights==null) Flights = 0; else Flights=Flights; if (Distance==null) Distance = 0; else Distance=Distance;</pre> <p>berserta mengubah tipe data yang berbeda dengan tipe data "string"</p> <p>Jumat,4 Nov 2022 membuat report di word yang dimulai identitas diri, menambahkan screenshot hasil data transformasi yang sudah dibuat, yang di dalamnya terdapat informasi logging sukses, screenshot 3 buah dashboard yang telah dibuat, dilengkapi dengan penjelasan cara penggunaanya (misal filter nya apa saja) dan juga insight dari masing - masing dashboard..</p>
<p>Minggu ke-13 07 - 11 Nov 2022</p>	<p>Pengerjaan Final project, Cycle meeting, dan Mentoring</p>	<p>pembelajaran dengan trainer kali ini digantikan pembelajaran asynchronous. membuat rangkuman big data mengenai Hadoop Ecosystems yang terdiri dari - What is Hadoop ? - Why Hadoop? - Hadoop Ecosystem -Why MapReduce? perbedaan apache spark vs hadoop map reduce terlihat apache spark dapat digunakan untuk pemrosesan batch dan hampir real-time. spark berpengaruh mampu menangani kebutuhan pemrosesan data apa pun. di sisi lain, map reduce kurang fiexsibel dibandingkan spark dalam hal melakukan berbagai pekejaan pemrosesan data, sementara mapreduce bisa menjadi salah satu alat pemrosesan batch terbaik yang tersedia. pembelajaran dengan trainer kali ini digantikan pembelajaran asynchronous membuat rangkuman big data mengenai Kafka & Spark yang terdiri dari - What is Apache Spark? - Why Apache spark? - What is Apache Kafka ? - Why Apache Kafka? pembelajaran dilakukan mandiri, mempelajari definisi dan kegunaan apache kafka apache Kafka adalah platform streaming acara yang digunakan untuk mengumpulkan, memproses, menyimpan, dan mengintegrasikan data dalam skala besar. Ini memiliki banyak kasus penggunaan termasuk logging terdistribusi. pemrosesan aliran, integrasi data, dan perpesanan pub/sub. kafka benar- benar membuat pemindahan data menjadi cepat pembelajaran live session dengan trainer, pembelajaran kedepannya menggunakan pyspark dan</p>

		<p>apache kafka. pembelajaran kali ini belajar hal - hal dasar (teori) mencakup spark sangat bagus & cepat untuk menjalankan program. beberapa perusahaan juga menggunakan spark. spark dibangun di sekitar satu konsep utama: Resilient Distributed Dataset (RDD). RDD sangat bagus jika ingin tetap menyimpan kumpulan data di memori dan mengaktifkan \$erangkaian kueri - ini berfungsi lebih baik daripada mengambil data dari disk setiap saat. komponen RDD. komponen spark (spark, MLLib, graphx)</p>
<p>Minggu ke-14 14 - 18 Nov 2022</p>		<p>senin pembelajaran dgn trainer diganti menjadi pembelajaran mandiri mengenai merangkum "Introduction to ML" hingga "Reducing Loss" dan hands-on ada pada "First Step with TF" . bisa memilih dari 2 case tersebut. adapun topik case nya latihan regresi linear dengan data sintesis, yang mengeksplorasi regresi linear dengan set data mainan dan regresi linear dengan set data nyata yang memandu memahami jenis analisis yang harus dilakukan pada set data sebenarnya selama mempelajari course ML, dengan topik descending into ML: linear regression. dapat dipahami model ini hanya menggunakan satu fitur. rabu penjelasan dari trainer : menjelaskan data split pada small data & big data di real casenya. cross validation akan berpengaruh pada akurasi data dan bisa melihat keseluruhan data. cross validation best practice untuk sampel data. setelah itu menggunakan modul seaborn untuk memprediksi data tips. mengubah data kategorikal menjadi numerik sehingga bisa diproses menjadi regresi. penjelasan activation functions (sigmoid) proses pembuatan model terdiri dari data preparation (EDA, Missing V, Duplicated, Balanced, Distribution) Features E (kolom/informasi/fitur yang akan digunakan) model selection (XGBoost, NN, Naive Bayes & SVM yg akan di eksplor. perlunya mengetahui karakteristik) Hyper Parameter turning (LearningRate & Fun Act menyesuaikan model yang dipilih juga kamis menemukan perbedaan lebih spesifik antar supervised learning: training data yang digunakan mencakup output yang diinginkan. membedakan mana data yang merupakan spam, mana yang bukan, pembelajarannya diawasi. unsupervised learning: training data yang digunakan tidak mencakup output yang diinginkan. lebih sulit untuk mengetahui mana yang merupakan pembelajaran yang baik dan tidak, contoh dari pembelajaran ini adalah clustering. Pembelajaran dengan tipe supervised learning adalah pembelajaran yang paling banyak dipelajari dan digunakan dalam algoritma machine learning, juga adalah pembelajaran yang paling dewasa. pembelajaran</p>

		<p>dengan supervision lebih mudah dilakukan dibandingkan tanpa supervision. Jumaat pembelajaran live session dilakukan trainer. pembahasan awal dimulai konsep ml : supervised learning ada classification terletak pada data diskrit (naive bayes, neural network, sv machine) dan regression terletak data kontinu (linear regression, logistic regression, polynomlal regression unsupervised learning ada clustering (k-means, DV SCAN, density based clustering). menjelaskan ruang lingkup AI (ML, natural language, Expert System, speech decognition dan vision). vision sangat berefek dan efisien jika diterapkan di real time. memperkenalkan hungging face (website yang spesifik dan ranahnya ke machine learning) keterkaitan AI dalam big data akan menghasilkan insight. pembahasan akhir mengenai: untuk proses ml pastinya ada data input, proses/algortma ml dan output selanjutnya hands on implementasi google collab pada classification naive bayes. mempersiapkan data dgn dataset nusax, hanya menerima alphabet dan numerik. menerapkan fungsi cleansing untuk kolom text</p>
<p>Minggu ke-15 21 - 25 Nov 2022</p>	<p>Pengerjaan Final project, Cycle meeting, dan Mentoring</p>	<p>Senin Tidak ada operasional hanya digantikan tugas dan zoom untuk menggantikan kelas pada hari senin Selasa pada topik pembelajaran machine learning, sering melakukan klasifikasi teks menggunakan python dan scikit-learn. dan terdapat algortma tingkat lanjut di Scikit Learn yang memiliki fungsi canggih seperti: metode ensemble: boosting, bagging, random forest, model voting dan averaging manipulasi fitur: pengurangan dimensi, pemilihan fitur, analisis fitur deteksi outlier: untuk mendeteksi outlier dan menolak noise pemilihan dan validasi model: validasi silang, penyesuaian hyperparameter, dan metrik</p> <p>Rabu pembelajaran dengan trainer, hal yang dipelajari text classification with TensorFlow Hub: Movie reviews(menunjukkan penerapan transfer dengan TensorFlow Hub) cara mencari perpotongan Y dan NLP Timeline Kamis menemukan wawasan mengenai model regresi linier dimana salah satu model machine learning yang paling sederhana. namun memiliki kompleksitas yang rendah namun optimal pada dataset yang memiliki hubungan linear. model linear regression ini juga merupakan representasi dari visualisasi data menggunakan diagram scatter plot. ada model regresi, ML akan memprediksi bilangan numerik berdasarkan atribut yang ada Jumaat unsupervised learning diawali mengkategorikan daya berdasarkan kemiripan jenis unsupervised yaitu clustering, k - means, DBSCAN, density based clustering. unsupervised learning biasanya ditujukan</p>

		mencari pola2 baru segmentasi market, behavior user yang mana proses mengolah datanya yang belum ada label. k-means (key number of cluster). k disini yang dimaksud k yang berada di dalam cluster. Konsep dasar dari K-Means adalah pencarian pusat cluster secara iteratif dan prinsip utama k-means adalah algoritma berbasis centroid
Minggu ke-16 28 Nov - 02 Des 2022	Live mentoring, Cycle meeting dan mengerjakan Final Project.	Senin pembelajaran dengan trainer yaitu handson DBSCAN, mengidentifikasi elbow-point, menjelaskan gambaan deep learning, karakteristik deep learning, hyperparameter Selasa memahami dampak Impact of Learning Rate pada Performa Neural Network. learning rate adalah hyperparameter yang mengontrol seberapa banyak perubahan model sebagai respons terhadap perkiraan kesalahan setiap kali bobot model diperbarui. learning rate mengontrol seberapa cepat model disesuaikan dengan masalah. learning rate yang lebih kecil memerlukan lebih banyak masa pelatihan mengingat perubahan yang lebih kecil yang dilakukan pada bobot setiap pembaruan, sedangkan learning rate yang lebih besar menghasilkan perubahan yang cepat dan memerlukan masa pelatihan yang lebih sedikit. Rabu eguralization of deep learning terdapat 2 hal : -drop out (yang berbentuk presentasi neuron yang akan di drop/dinonaktifkan) -early stopping(terjadi sebelum terjadi overfit)berlaku untuk arsitektur machine learning arsitektur pada deep learning memiliki di desain sesuai problem yang akan diselesaikan umumnya ada 2 arsitektur CNN(convolutional neural network) dan RNN(recurrent neural network) CNN(convolutional neural network) sisi input - bisa mengelola data apapun seperti gambar. output yang dihasilkan kelas/kategori. implementasinya pada image classification, object detection RNN(recurrent neural network)melabel data sekuensial seperti teks. output yang dihasilkan kelas sekuensial. implementasinya pada NCP, sentimen Kamis selain learning rate yang relate dengan training algoritma, melainkan ada momentum, number of epochs dan batch size. metode yang digunakan untuk mengetahui hyperparameter: ada manual search grid search, random search dan bayesian optimization Jumaat pembelajaran dengan trainer tentang fundamental deep learning, meneruskan pembelajaran sebelumnya dengan image classification. salah satu mencegah overfitting bisa menggunakan data augmentation dan dropout regularization. mempelajari teori & handson LSTM, implementasi RNN(ada training tensorflow, penggunaan early stopping) melihat perbedaan RNN & LSTM yang terdapat pada hidden layer.

Minggu ke-17 05 - 09 Des 2022	Live mentoring, Cycle meeting dan mengerjakan Final Project.	Pada Hari senin kami diberi materi dan penjelasan tentang pengerjaan Tugas project akhir yang telah di berikan ke kami agar pengerjaan bisa dikerjakan dengan maksimal pada sesi ini memahami alur dari pengerj aan project akhir data secara end to end selanjutnya pada sesi kali ini sedang dalam pengerjaan project akhir dimana membuat model dan dan data preperation pada sesi kali ini sedang dalam pengerjaan project akhir dimana membuat build model di google colab atau di sebut phyton pada sesi kali ini sedang dalam pengerjaan project akhir dimana membuat preditck data dan menyiapkan model data nya di data properation
Minggu ke-18 12 - 16 Des 2022	Live mentoring, Cycle meeting dan mengerjakan Final Project.	pada sesi kali ini sedang dalam pengerjaan project akhir diamana membuat model dan proses data pada sesi kali ini sedang dalam pengerjaan project akhir dimana membuat model dan visualisassi data sesi kali ini sedang dalam pengerjaan project akhir dimana mengecek lagi persiapan data menggunakan phyton dan visualisasi menggunakan tableu pada sesi kali ini sedang dalam pengerjaan project akhir dimana membuat modelnya lagi dan mngesekusi dari data tree ny pada sesi kali ini sedang dalam pengerjaan project akhir dimana membuat model dan proses visualisaasi tersebut agar mudah dipahami
Minggu ke-19 19 - 23 Des 2022	Live mentoring, Cycle meeting dan mengerjakan Final Project.	pada sesi kali ini sedang dalam pengerjaan project akhir dimana membuat laporan dan melanjutkan visualisasi model data pada sesi kali ini sedang dalam pengerjaan project akhir dimana membuat model dan dasbord nya di tableu pada sesi kali ini sedang dalam pengerjaan project akhir dimana lmembuat laporan dan video pembahasan pada sesi kali ini sedang dalam pengerjaan project akhir dimana lmembuat laporan dan video pembahasan ada sesi kali ini sedang dalam pengerjaan project akhir dimana mengecek kembali dari awal dari data preparation dan modeling datanya
Minggu ke-20 26 - 30 Des 2022	Live mentoring, Cycle meeting dan mengerjakan Final Project.	selanjutnya pada sesi kali ini sedang dalam pengerjaan project akhir dimana kami mengumpulkan data hasil akhir dari project yang kita buat seperti video, hasil laporan dan hasil visulisasi dan modelingg data tersebut Pada sesi ini tidak ada kegiatan setelah kita mengumpulkan project akhir kemarin Pada sesi ini adanya sesi dimana kelulusan dari msib celerates Pada sesi selanjutnya sampai hari terakhir tidak ada kegiatan

BAB VIII

Lampiran D. Surat Pernyataan Kerahasiaan



PT Mitra Talenta Grup

The Manhattan Square Building, Mid Tower, 12th Floor,
Jl. T.B. Simatupang Kav 1-S, Cilandak Timur, Jakarta Selatan

Celerates.com

contact@celerates.com
021 8064 1000

SURAT PERNYATAAN KERAHASIAAN

No. 001/002/KM-MTG/XII/2022

Yang bertanda tangan di bawah ini,

Nama : Muhamad Risyad Ganis

Jabatan di Perusahaan : Chief Operating Officer

Nama Perusahaan : PT Mitra Talenta Grup (Celerates)

Melalui surat ini, kami menyatakan bahwa informasi yang dibutuhkan mahasiswa Kampus Merdeka Batch 3 Studi Independen Celerates Acceleration Program (Big Data & Business Intelligence) di PT Mitra Talenta Grup (Celerates) untuk keperluan Laporan Akhir Mahasiswa bersifat RAHASIA. Informasi yang dimaksud adalah:

1. Informasi mencakup struktur organisasi, nama karyawan, unit atau divisi yang ada di dalam PT Mitra Talenta Grup (Celerates).
2. Informasi mencakup posisi dan deskripsi pekerjaan yang ada di dalam PT Mitra Talenta Grup (Celerates).

Atas keputusan tersebut, surat ini dibuat sebagai pengganti halaman laporan akhir mahasiswa yang membutuhkan informasi yang kami nyatakan sebagai informasi rahasia.

Jakarta, 20 Desember 2022
PT Mitra Talenta Grup (Celerates)

Muhamad Risyad Ganis
Chief Operating Officer



SERTIFIKAT MSIB

Diberikan Kepada :

WAHYU CITRA PRATAMA

ID Kegiatan : 3801280 - Universitas Esa Unggul - TEKNIK INFORMATIKA

Sebagai :

Peserta MSIB Angkatan 3

Telah berhasil menyelesaikan tugasnya di **PT Mitra Talenta Grup (Celerates)** dalam **program Magang & Studi Independen Bersertifikat (MSIB) Kampus Merdeka** dengan kegiatan **Studi Independen, Big Data & Business Intelligence** yang diselenggarakan pada **tanggal 18 Agustus - 31 Desember 2022.**

Ketua Program Mitra,
PT Mitra Talenta Grup (Celerates)

Muhamad Risyad Ganis

CAPAIAN PEMBELAJARAN PROGRAM

No	Kompetensi	Definisi Kompetensi	Jam	Nilai Capaian	Deskripsi Nilai Capaian
1	Database & sql	- Mampu melakukan query terhadap data terstruktur - Mampu melakukan Exploratory Data Analysis untuk mengambil insight dari data mentah - Mampu melakukan fungsi Select, Join, Agregat, dan CTE	80 Jam	70	Penilaian akan berfokus pada keahlian peserta dalam melakukan Query yang dinilai dari tugas yang diberikan kepada peserta, selain itu pengajar akan menilai keaktifan dan ketekunan peserta selama pengajaran synchronous
2	Python for Data Analytics	- Mampu melakukan analisis data dengan pendekatan programming python - Mampu melakukan fungsi aritmetis dan statistik terhadap data menggunakan python - Mampu melakukan Exploratory Data Analytics dengan berbagai teknik	80 Jam	75	Penilaian akan dilakukan melalui: - Nilai peserta dari pengerjaan soal tugas - Nilai peserta dari kuis - Keaktifan dan ketekunan peserta selama pelaksanaan synchronous
3	Data Warehouse	- Mampu memahami konsep dan karakteristik Data Warehouse - Mampu memahami dimensi(Measure,dan fact Tabel - Mampu memahami serta membuat skema pada Data Warehouse - Mampu membedakan OLTP dan OLAP	125 Jam	75	Penilaian akan dilakukan melalui: - Nilai peserta dalam memahami konsep maupun pengerjaan datawarehouse - Nilai peserta dalam memahami/membuat ERD dan Schema DWH - Nilai Peserta dalam menentukan data OLTP dan OLAP - Nilai presentasi tugas akhir
4	ETL (Perintah)	- Mampu melakukan instalasi serta konfigurasi Perintah Data Integration - Mampu membuat script/program maupun mengimpor sebuah repository - Mampu memahami perspektif Perintah Data Integration - Mampu melakukan serta mengembangkan sebuah transformasi serta job	125 Jam	80	Penilaian akan dilakukan melalui: - Nilai peserta dalam memahami perspektif Perintah Data Integration sebagai ETL tools - Nilai peserta dalam melakukan proses development transformasi maupun job - Nilai peserta dalam melakukan handling error
5	Data Visualization (Tableau & Power BI)	- Mampu memahami fungsi Data Visualization - Mampu memahami terminology pada Tableau/Power BI - Mampu melakukan pemanggilan data source - Mampu membedakan tipe data,dimensi,dan measure - Mampu membuat basic Chart - Mampu membuat Dashboard dan insight	80 Jam	80	Penilaian akan dilakukan melalui: - Nilai terhadap pemilihan topik/case sumber data - Nilai dalam penyusunan layout dashboard - Nilai dalam menentukan metrics dashboard - Nilai dalam menarik kesimpulan untuk menghasilkan sebuah insight dari dashboard - Nilai terhadap cara presentasi dan kelengkapan power point - Keaktifan dan ketekunan
6	Big Data & Cloud Technology	- Mampu memahami apa itu Big Data - Mampu memahami perbedaan hads, SaaS, PaaS - Mampu memahami fungsi Hadoop Ecosystem - Mampu menggunakan Spark, Kafka, HDFS, dan Airflow - Mampu menggunakan Google Cloud Platform dan memahami fitur-fitur didalamnya - Mampu melakukan data processing secara batch maupun streaming - Mampu melakukan data orchestration	125 Jam	85	Penilaian akan berfokus pada keahlian peserta dalam melakukan Query yang dinilai dari tugas yang diberikan kepada peserta, selain itu pengajar akan menilai keaktifan dan ketekunan peserta selama pengajaran synchronous
7	Statistic	- Mampu memahami dasar-dasar statistika - Mampu memahami penggunaan ilmu statistika dalam analisa data - Mampu melakukan validasi data - Mampu memahami kesalahan statistik - Mampu melakukan sampling dengan metode yang benar - Mampu melakukan uji hipotesis dengan metode statistik	35 Jam	70	Penilaian dilakukan berdasarkan dokumentasi tugas akhir individu.
8	Machine Learning	- Mampu memahami Bayesian Probability - Mampu memahami teknik-teknik Machine Learning (Random Forest, Naive Bayes, dll) - Mampu memahami teknik ML terhadap data terstruktur - Mampu melakukan teknik ML (klasifikasi, regresi) menggunakan Python	80 Jam	90	Penilaian dilakukan pada saat project akhir materi Machine Learning dengan dokumentasi dan presentasi peserta.
9	Deep Learning	- Mampu memahami perbedaan antara Machine Learning dengan Deep Learning - Mampu memahami teknik-teknik Deep Learning (CNN, RNN, LSTM) - Mampu menggunakan teknik-teknik Deep Learning untuk melakukan klasifikasi terhadap data tidak ter - Mampu menggunakan Tensorflow dalam melakukan praktik Deep Learning	80 Jam	80	Penilaian dilakukan pada saat project akhir materi Deep Learning dengan dokumentasi dan presentasi peserta.
10	Final Project	- Mampu memahami alur pengerjaan analisis data secara end-to-end - Mampu melakukan cleaning data - Mampu membuat model machine learning - Mampu membuat dashboard analisa - Mampu melakukan manajemen waktu - Mampu menentukan analisa mandiri terkait data - Mampu membuat portfolo pribadi	90 Jam	85	Penilaian akan dilakukan melalui: - Nilai peserta dalam memahami sumber data tugas akhir - Nilai peserta dalam memahami studi kasus tugas akhir - Nilai peserta berdasarkan pengolahan sumber data tugas akhir - Nilai peserta dalam memahami algoritma machine learning yang digunakan pada tugas akhir - Nilai peserta berdasarkan akurasi prediksi terhadap data testing - Nilai peserta dalam penyusunan layout dashboard tugas akhir - Nilai peserta dalam menentukan metrics dashboard tugas akhir - Nilai peserta dalam mengaitkan dokumentasi tugas akhir - Nilai peserta berdasarkan ketekunan dan ketekunan tugas akhir yang telah dibuat
Nilai Rata-Rata				81	PREDIKAT : A