LAPORAN TENGAH/AKHIR

MAGANG & STUDI INDEPENDEN BERSERTIFIKAT

EDUCATION: STUDENT PERFORMANCE

Di PT Digitalisasi Pemuda Indonesia

Diajukan untuk memenuhi persyaratan kelulusan

Program MSIB MBKM

oleh :

FORDINANDHALOMOAN PASARIBU / 188160061

TEKNIK INFORMATIKA

UNIVERSITAS MEDAN AREA

2021 / 2022

Lembar Pengesahan Teknik Informatika

EDUCATION: STUDENT PERFORMANCE

Di PT Digitalisasi Pemuda Indonesia

oleh :

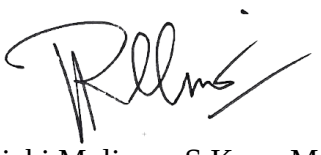
FORDINAND HALOMOAN PASARIBU / 188160061

disetujui dan disahkan sebagai

Laporan Magang atau Studi Independen Bersertifikat Kampus Merdeka

Medan, 17 Januari 2022

Pembimbing Magang atau Studi Independen Teknik Informatika Universitas Medan Area.



Rizki Muliono, S.Kom, M.Kom

NIP: 0109038902

Lembar Pengesahan

EDUCATION: STUDENT PERFORMANCE

Di PT Digitalisasi Pemuda Indonesia

oleh :

FORDINAND HALOMOAN PASARIBU / 188160061

disetujui dan disahkan sebagai

Laporan Magang atau Studi Independen Bersertifikat Kampus Merdeka

Bandung, 17 Januari 2022

Mentor

Ignasius Frans D.S.T.N, S.Kom

Abstraksi

Laporan ini bertujuan untuk memahami nilai ujian siswa yang berpengaruh pada Jenis Kelamin, Suku/Ras, Tingkat Pendidikan Orang Tua, Makan Siang, dan Kursus persiapan ujian. Kemudian Membuat model *Machine Learning* untuk memprediksi nilai siswa Analisis ini menggunakan variabel independen yaitu Jenis Kelamin, Suku/Ras, Tingkat Pendidikan Orang Tua, Makan Siang, dan Kursus persiapan ujian. Variabel dependennya yaitu keterangan lulus atau tidak lulus. Data yang digunakan merupakan data yang sudah di sediakan di situs Kaggle. Adapun algoritma yang diterapkan ialah *KNearest Neighbors* (KNN), *Support Vector Machine* (SVM), *Multilayer Perceptron* (MLP). Alur pemodelannya yaitu: Pengumpulan Data, *Exploratory Data Analysis* (EDA), *Data Preprocessing*, dan Evaluasi Model. Hasil dari yang sudah dilakukan oleh penulis menunjukkan bahwa penerapan algoritma SVM memiliki akurasi sebesar 98%, dan variabel seperti makan siang sangat berpengaruh terhadap nilai ujian siswa.

Kata Kunci: Nilai ujian, *Machine Learning*, KNN, SVM, MLP

Kata Pengantar

Puji dan syukur saya sampaikan kepada Tuhan Yang Maha Esa karena telah memberikan rahmat dan karunia-Nya kepada saya, sehingga berhasil menyelesaikan laporan ini tepat pada waktunya tentang Tugas Proyek Tengah/Akhir Magang dan Studi Independen Bersertifikat (MSIB) dengan judul *Education: Student Performance*. Laporan ini adalah salah satu syarat kelulusan peserta MSIB dalam program Kampus Merdeka yang dibuat oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia. Saya telah banyak memperoleh bantuan dan bimbingan dari berbagai pihak, maka pada kesempatan ini saya menyampaikan ucapan terima kasih kepada :

1. Kak Aditya Soleh, Selaku Officer at Digital Skola.
2. Kak Agil Haykal, Kak Romansya Setyo, dan Pak Ganjar Alfian , Selaku Tutor Digital Skola.
3. Kak Shavira Tisna dan Kak Qoriyana Nurselvi, Selaku Representative at Digital Skola.
4. Kepada mentor kelompok Fantastic Five yaitu Kak Ignasius Frans, dan teman-teman: Denta Winardi, Rahmatul Fajri, Nur Afni, dan Zulva Amaliya.

Terima kasih atas bantuan, bimbingan, dukungan, fasilitas yang telah diberikan kepada saya. Semoga Tuhan Yang Maha Esa membalas semua kebaikan yang telah diberikan kepada saya. Saya menyadari bahwa laporan ini jauh dari kata kesempurnaan, oleh karena itu kritikan dan saran yang bersifat membangun selalu saya harapkan demi kesempurnaan laporan ini. Akhir kata saya sampaikan terima kasih kepada pembaca yang sekiranya telah meluangkan waktunya untuk membaca laporan ini seraya memajukan selangkah lagi pengetahuan tentang isi laporan ini.

Medan, 17 Januari 2022

Fordinand Halomoan Pasaribu

NIM: 188160061

Daftar Isi

Daftar Gambar

Bab I Pendahuluan

I.1. Latar belakang

Dalam rangka menyiapkan mahasiswa menghadapi perubahan sosial, budaya, dunia kerja dan kemajuan teknologi yang pesat, kompetensi mahasiswa harus disiapkan untuk lebih gayut dengan kebutuhan zaman. *Link and match* tidak saja dengan dunia industri dan dunia kerja tetapi juga dengan masa depan yang berubah dengan cepat. Perguruan Tinggi dituntut untuk dapat merancang dan melaksanakan proses pembelajaran yang inovatif agar mahasiswa dapat meraih capaian pembelajaran mencakup aspek sikap, pengetahuan, dan keterampilan secara optimal dan selalu relevan.

Kampus Merdeka diharapkan dapat menjadi jawaban atas tuntutan tersebut. Kampus Merdeka merupakan wujud pembelajaran di perguruan tinggi yang otonom dan fleksibel sehingga tercipta kultur belajar yang inovatif, tidak mengekang, dan sesuai dengan kebutuhan mahasiswa. Proses pembelajaran dalam Kampus Merdeka merupakan salah satu perwujudan pembelajaran yang berpusat pada mahasiswa (*student centered learning*) yang sangat esensial. Pembelajaran dalam Kampus Merdeka memberikan tantangan dan kesempatan untuk pengembangan inovasi, kreativitas, kapasitas, kepribadian, dan kebutuhan mahasiswa, serta mengembangkan kemandirian dalam mencari dan menemukan pengetahuan melalui kenyataan dan dinamika lapangan seperti persyaratan kemampuan, permasalahan riil, interaksi sosial, kolaborasi, manajemen diri, tuntutan kinerja, target dan pencapaiannya.

Digital Skola merupakan perusahaan yang baru didirikan oleh tiga anak bangsa yang bertujuan mendorong pertumbuhan talenta digital di Indonesia dengan menghadirkan pusat edukasi nonformal secara daring mengenai kecakapan digital yaitu *data scientist, data engineer* dan *digital marketing*. Digital Skola berkomitmen mewujudkan pelatihan *skill digital* yang inklusif dengan biaya terjangkau, waktu fleksibel, pengajar kredibel dan kurikulum sesuai kebutuhan industri. Dalam pelaksanaan pelatihannya, Digital Skola bekerja sama dengan tutor-tutor yang merupakan praktisi di bidang masing-masing.

Dari segi penyusunan kurikulum serta metode belajar, Digital Skola memberikan jaminan bahwa semua peserta dari latar belakang apapun akan mampu menguasai ilmu yang diajarkan. Kurikulum pun disusun dengan cermat sesuai kebutuhan industri agar lulusan pelatihan dari Digital Skola menguasi kemampuan untuk langsung dapat bekerja. Berikut adalah visi & misi Digital Skola :

1. Visi: Mengembangkan wawasan dan kemampuan angkatan kerja di Indonesia untuk menjadi generasi.

2. Misi: Mengadakan pelatihan informal keahlian digital dengan biaya terjangkau, waktu fleksibel, pengajar kredibel dan kurikulum sesuai kebutuhan industri.

I.2. Lingkup

- Nama Kegiatan.
- Latar Belakang.
- Tujuan Kegiatan.
- Jadwal dan Waktu Pelaksanaan.
- Peserta.
- Penyelenggara.

I.3. Tujuan

Data Science meliputi pembelajaran individu dan implementasi proyek dalam bentuk tim. Pada pembelajaran individu, setiap peserta akan mengikuti kelas secara *interactive live-online meeting* dimana peserta akan mendengar secara langsung serta dapat berkonsultasi dengan *expert* serta pembimbing terkait materi yang dipelajarinya. Hasil yang akan diperoleh, yaitu:

1. Mampu memahami dasar-dasar statistika yang digunakan pada domain *data science*.
2. Mampu menguasai bahasa pemrograman Python termasuk *libraries* yang digunakan.
3. Mampu mengerjakan *dataset* yang diberikan serta membuat *modelling* nya.
4. Mampu membuat model *machine learning* sesuai kebutuhan.
5. Mampu membuat visualisasi data secara efektif.
6. Memahami materi yang terkait dengan *Analytics*, yang meliputi metodologi *data science*, Numpy, *dataframe*, statistik, visualisasi data, dan *business intelligence*.

Bab II Digital Skola

II.1. Struktur Organisasi

Tim Digital Skola

1. Aditya Soleh (Chief Executive Officer).
2. Zico Alaia Akbar (Chief Operating Officer).
3. Stephanie Octavia (Chief Business Development Officer).
4. Indah Salimin (Senior Copywriter).
5. Ismaya R. P. (Graphic Designer).
6. Qoriyana Nurselvi (Product Officer).
7. Neti Rosmayanti (Customer Relations Officer).
8. Faliha Ishma (Product Officer).
9. Siti Meiranti Nabilah (Account Executive).
10. Adellia Anggun Trisnawati (Customer Relations Officer).
11. Shavira Tisna (Representative).

Tutor Digital Skola

1. Agil Haykal (Analytics).
2. Mohammad Aprialdi (Data Scientist).
3. Faldi Sulistiawan (Data Scientist).
4. Farhan Reza (Analytics).
5. Muhammad Farid (Data Analyst).
6. Romansya Setyo (Business Intelligence Analyst).
7. Calvin Nixcon (Data Analyst).

8. Addinul Masri (Data Analyst).
9. Dr. Eng. Ganjar Alfian (Assistant Professor at one of the campus in Korea).
10. Ari Sulistiyo Prabowo (Senior Data Analyst).
11. Anastasia Yoan (Digital Marketing Specialist).
12. Claudia Natasya (Senior Sales & Marketing).
13. Gabriella Ovina (Head of Digital Minshare).
14. Irene Bergosa (Internet Marketing Lead).
15. Jovanka Audria (Content Specialist).
16. Laurensia Octavia (Founder Go Digital Consultant).
17. Deasy Natalia (Chief Operations Officer).
18. Salman Al Farisiy (Head of Marketing Communications).
19. Sherly Fusin (Senior Digital & Media Planning Manager).
20. Jaka Darmawan (SEO Specialist).
21. Nana Kusuma (Senior SEM).
22. Shona Firdaus (Performance Marketing Specialist).
23. Muhammad Haekal S.A (Senior Digital Marketing).
24. Ali Firdaus Chifari (Data Engineering).
25. Dimasta Juniar (Data Engineering).
26. Rio Harapan P (Big Data Engineering).
27. Rizki Dermawan (Data Engineering).
28. Ignes Nathania Widjaja (Product Designer).
29. Fahmi Kusuma Aji (UX Designer).
30. Areta Selena Khrista (UX Designer).

31. Mohamad Arief Bagusprastyo (UI/UX Designer).
32. Hary Nugraha (Software Engineering).
33. Dedi Ananto (Full-Stack Web Developer).

II.2. Lingkup Pekerjaan

Tim Digital Skola

1. Aditya Soleh (Chief Executive Officer).
2. Zico Alaia Akbar (Chief Operating Officer).
3. Qoriyana Nurselvi (Product Officer).
4. Shavira Tisna (Representative).

Tutor Program MSIB Data Science Digital Skola

1. Agil Haykal (Analytics).
2. Farhan Reza (Analytics).
3. Romansya Setyo (Business Intelligence Analyst).
4. Dr. Eng. Ganjar Alfian (Assistant Professor at one of the campus in Korea).

II.3. Deskripsi Pekerjaan

2.3.1. Nama Kegiatan

Dataset & Modelling Project

2.3.2. Latar Belakang

Machine Learning adalah metode yang digunakan untuk membuat program yang bisa belajar dari data. Berbeda dengan program komputer biasa yang statis, program *machine learning* adalah program yang dirancang untuk mampu belajar sendiri. Cara belajar program *machine learning* mengikuti cara belajar manusia, yakni belajar dari contoh-contoh. *Machine learning* akan mempelajari pola dari contoh-contoh yang dianalisa, untuk menentukan jawaban dari pertanyaan-pertanyaan berikutnya. Memang tidak semua masalah bisa dipecahkan dengan program *machine learning*. Namun, seringkali algoritma yang sifatnya kompleks, ternyata bisa dipecahkan dengan sangat simpel oleh *machine learning*. Adapun alur kerja *Machine Learning* yaitu:

1. **Perencanaan Model dan Pengumpulan data**

   Tentukan tujuan dari pembuatan *machine learning* dengan penguraian masalah yang ingin diselesaikan. Kemudian kumpulkan dataset yang relevan dengan permasalahan dan tujuan pemodelan. *Dataset* akan memberikan kontribusi besar terhadap akurasi model yang akan dibuat.

2. **Persiapan Data**

   - *Training* data yaitu data yang akan digunakan mesin untuk belajar.
   - *Testing* data yaitu data yang digunakan untuk prediksi.
   - *Validation* data yaitu data bentuk validasi model hasil belajar.

3. **Training Model**

   Gunakan algoritma yang sesuai dengan kebutuhan dan masalah yang ingin diselesaikan. Cari model yang terbaik dengan menggunakan *training* data. Pastikan fitur-fitur yang digunakan sebagai indikator sesuai dengan tujuan pemodelan.

4. **Evaluasi Model**

   Gunakan metrik evaluasi untuk mengukur performa model yang dihasilkan. Kemudian, gunakan *validation* data untuk merepresentasikan performa model saat digunakan pada data yang belum dipelajari, sehingga hasilnya dapat digunakan untuk proses *tuning*.

5. **Parameter Tuning**

   Lakukan *tuning* parameter untuk meningkatkan performa model yang dihasilkan. Contoh parameter yang dapat disesuaikan: jumlah tahapan *training*, *learning rate*, dan lain-lain. Kemudian,lakukan evaluasi kembali pada model yang sudah di *tuning*.

6. **Lakukan Prediksi**

   Ketika model yang diharapkan sudah sesuai dengan kebutuhan maka Langkah selanjutnya adalah melakukan uji prediksi menggunakan *testing* data. Tujuannya adalah mendapatkan gambaran yang lebih baik mengenai performa model ketika digunakan untuk memprediksi data di masa depan.

2.3.3. Tujuan Kegiatan

Adapun tujuan kegiatan *dataset & modelling project* untuk :

1. Prediksi yaitu melakukan prediksi berupa nilai, probabilitas maupun data dan kemudian merekomendasikan hasilnya untuk digunakan sebagai alat bantu pengambil keputusan maupun secara langsung digunakan secara otomatis oleh sistem.

2. Deskripsi yaitu menampilkan pola data untuk dianalisa dan penemuan masalah.

2.3.4. Jadwal dan Waktu Pelaksanaan

Tanggal *:* 23 Agustus 2021 – 28 Januari 2022

Jadwal Kelas :

- Selasa: jam 19.15 – 21.15 WIB
- Kamis: jam 19.15 – 21.15 WIB
- Sabtu: jam 13.00 – 15.00 WIB

Tempat : Online Via Zoom.

2.3.5. Peserta

Jumlah peserta kegiatan kelas SIB B sekitar 49 orang, dengan didominasi oleh mahasiswa dari program studi Informatika, Sistem Informasi, Ilmu Komputer, Geofisika, Geologi Elektro, Matematika, Akuntansi, Statistika, dan Manajemen.

2.3.6. Penyelenggara

Penyelenggara dari kegiatan ini adalah PT Digitalisasi Pemuda Indonesia (Digital Skola).

II.4. Jadwal Magang dan Studi Independen Bersertifikat

Sesi pembelajaran online sebanyak 60.

Pembelajaran dan sesi bimbingan sebanyak 600 jam.

Tanggal : 23 Agustus 2021 – 28 Januari 2022

Jadwal Kelas	
SIB A	SIB B
Senin : 19.15 - 21.15 WIB	Selasa: 19.15 - 21.15 WIB
Rabu : 19.15 - 21.15 WIB	Kamis: 19.15 - 21.15 WIB
Jumat : 19.15 - 21.15 WIB	Sabtu: 13.00 - 15.00 WIB

Tempat : Online Via Zoom.

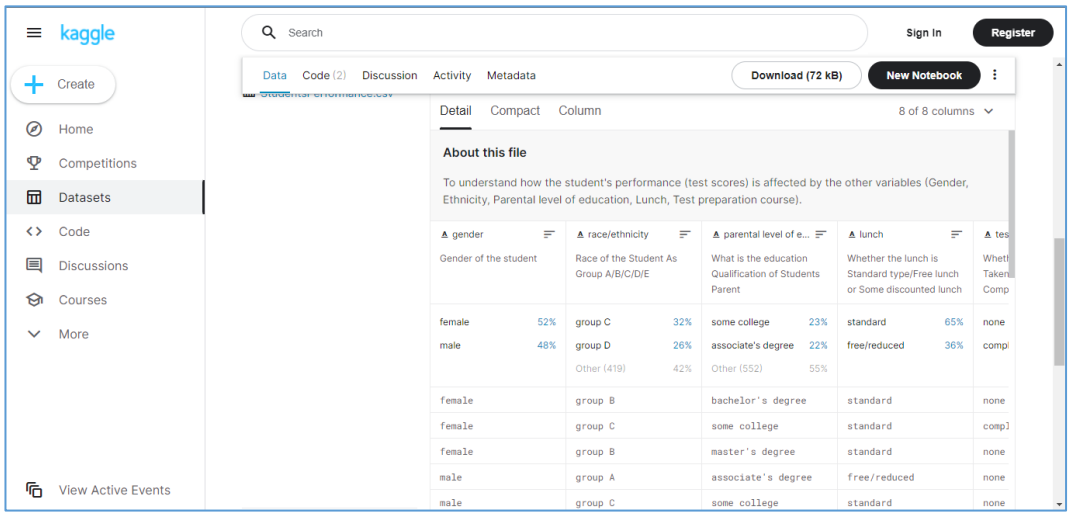
Bab III *Education: Student Performance*

III.1. Deskripsi/Konteks

Bagaimana memahami kinerja siswa(nilai ujian) yang dipengaruhi oleh variabel lain (Jenis Kelamin, Suku/Ras, Tingkat Pendidikan Orang Tua, Makan Siang, dan Kursus persiapan ujian). Adapun nilai-nilai ujian yang diprediksi dapat dipengaruhi oleh variabel lain meliputi Nilai Matematika, Nilai Membaca, dan Nilai Menulis. Kemudian Membuat beberapa model *machine learning* untuk memprediksi nilai siswa. Adapun dataset yang penulis gunakan dalam MSIB ini adalah berasal dari situs https://www.kaggle.com/adithyabshetty100/student-performance

III.2. Proses Model *Machine Learning*

3.2.1 Pengumpulan Data

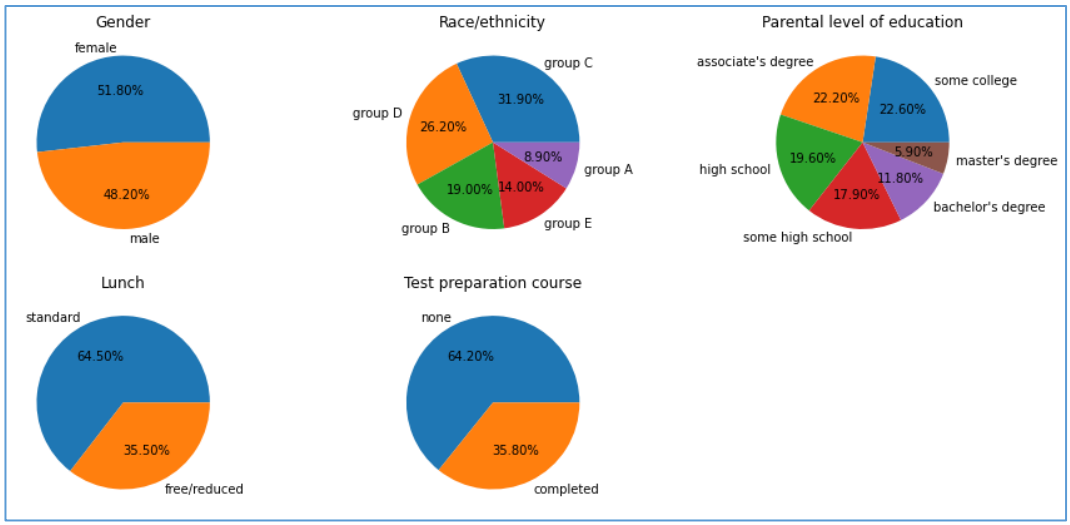


Gambar 1 Tampilan *dataset student performance*

Langkah pertama dalam pembuatan *machine learning* yaitu mengumpulkan data. Semakin banyak dan semakin baik kualitas data yang kita punya, performa *machine learning* yang kita buat akan semakin baik. Ada beberapa metode dalam mengumpulkan data seperti *web scrapping*, *data mining*, atau dari situs website yang sudah di sediakan seperti kaggle.com, data.go.id dan lain sebagainya. Adapun data yang digunakan dalam *project* MSIB ialah melalui situs kaggle.

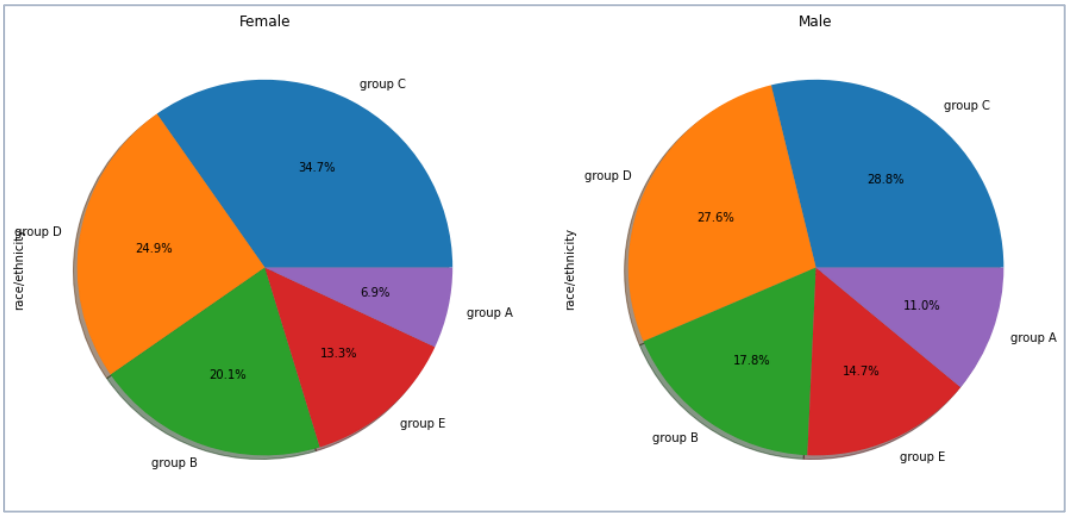
3.2.2 *Exploratory Data Analysis* (EDA)

EDA atau dikenal pula dengan analisis data eksploratif merupakan pendekatan analisis untuk suatu data guna membuat gambaran keseluruhan (*summary*) data sehingga mudah untuk dipahami. EDA memungkinkan analisis memahami isi data yang digunakan, mulai dari distribusi, frekuensi, korelasi dan lainnya. Dalam prakteknya, *curiousity* sangat penting dalam proses ini, pemahaman konteks data juga diperhatikan, karena akan menjawab masalah masalah dasar.



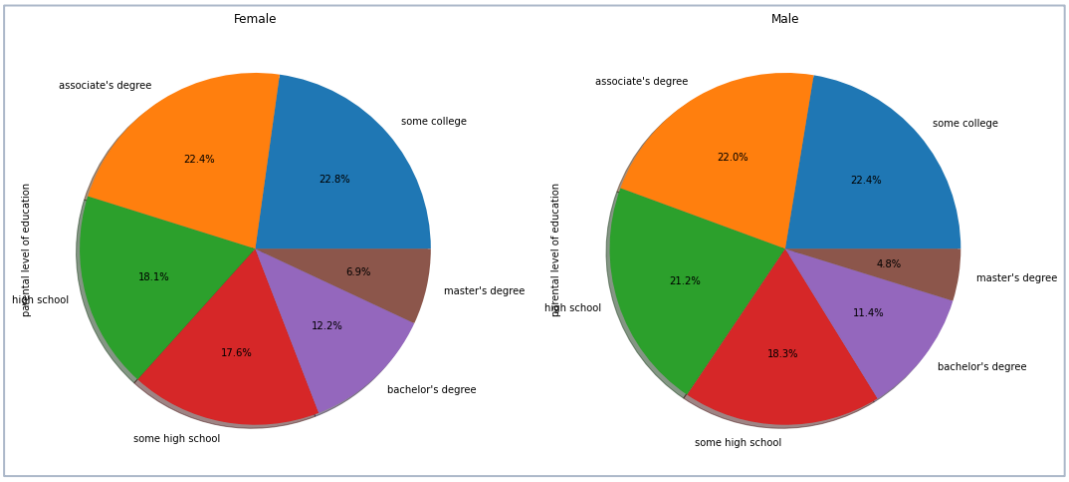
Gambar 2 Persentase data per kolom

Pada gambar 2 kita dapat melihat selisih perbandingan antar data pada tiap – tiap kolom, yaitu pada kolom *gender* didominasi oleh *female*, kolom *race/ethnicity* didominasi oleh *group C*, kolom *parental level of education* didominasi oleh *Some College* dan memiliki selisih 0,40% dengan *associate's degree*, kemudian kolom lunch didominasi oleh *standard*, dan terakhir kolom *test treparation course* didominasi oleh *none*(tidak mengikuti kursus).



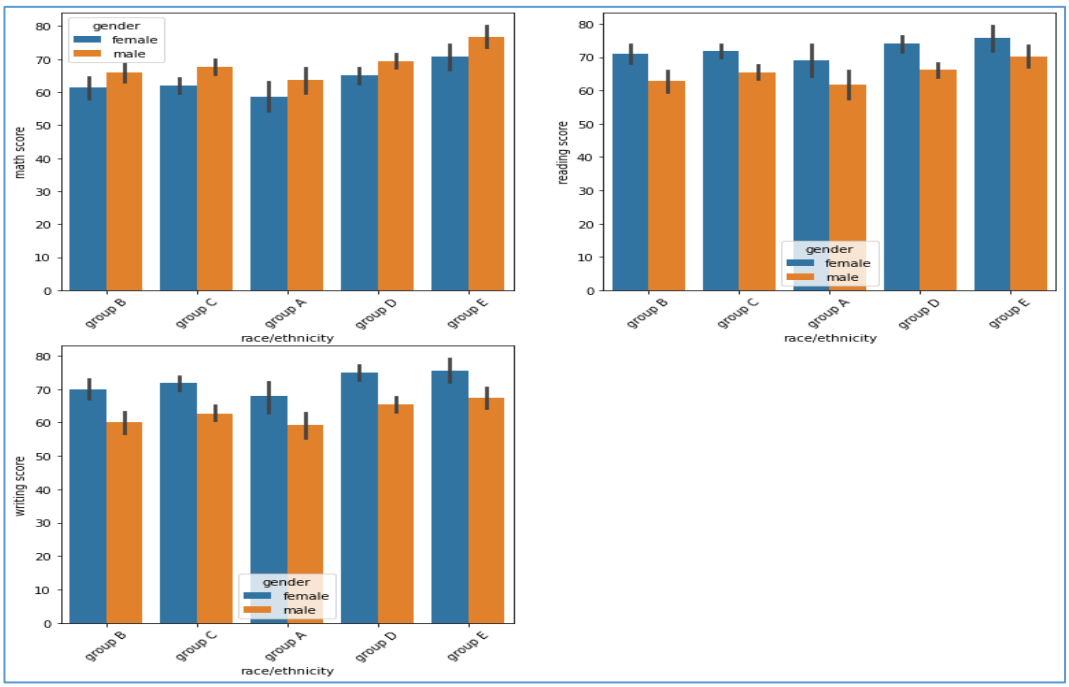
Gambar 3 *Race/Ethnicity* berdasarkan *Gender*

Pada gambar 3 bertujuan untuk membandingkan persentase *gender female* dan *male* terhadap *race/ethnicity*, dan hasil menunjukkan bahwa *group C* pada *gender female* lebih unggul sebesar 34,7%, kemudian pada *group D* terhadap *gender male* lebih unggul sebesar 27,6%, dan terakhir *group B* didominasi oleh *gender female* kembali dengan sebesar 20,1%.



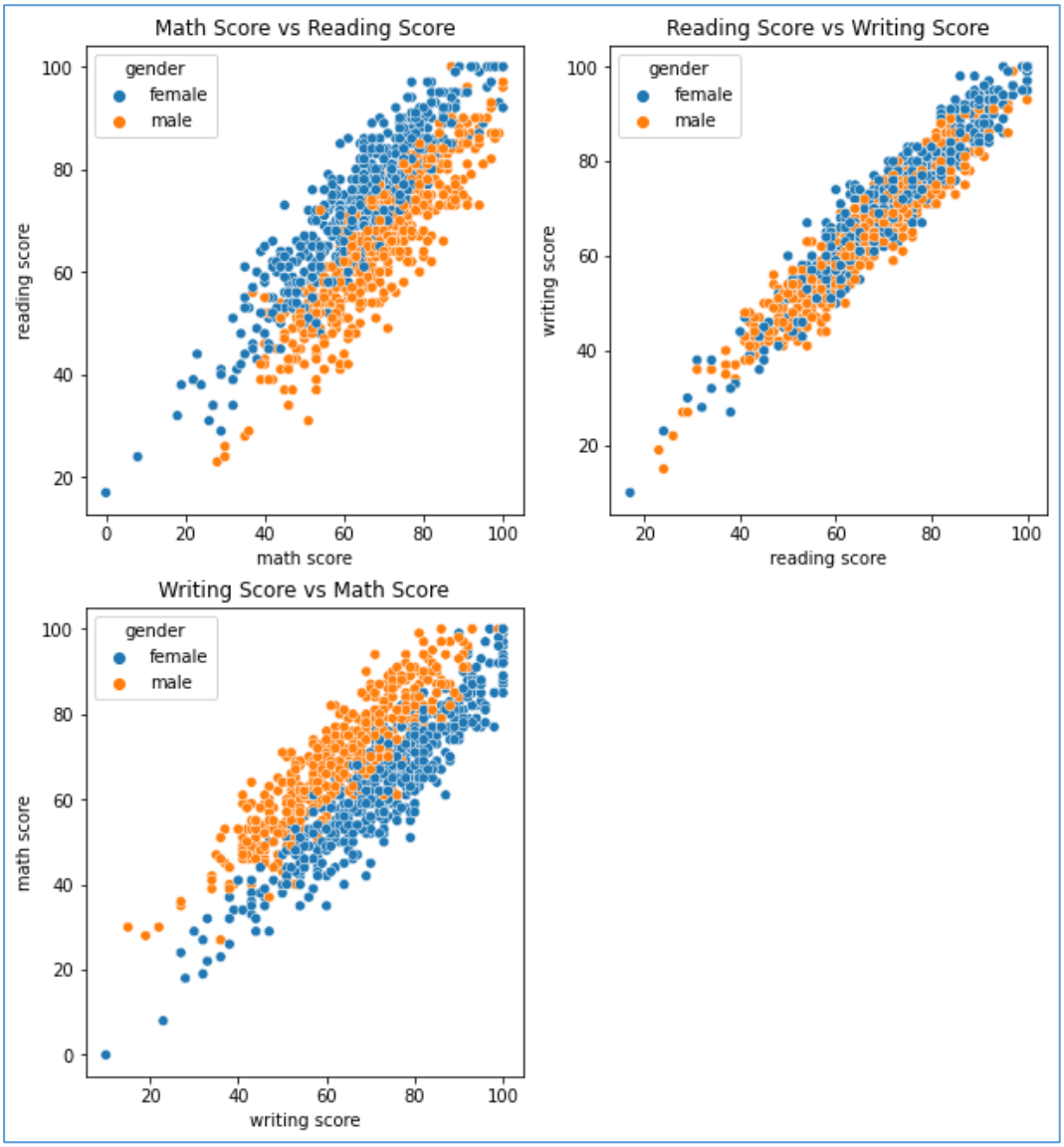
Gambar 4 *Parental level of education* berdasarkan *Gender*

Pada gambar 4 terlihat bahwa perbandingan *female* dengan *male* pada *associate's degree* memiliki selisih sebesar 0,4%, kemudian pada *Some College* juga selisih sebesar 0,4%, dan terakhir *high school* pada *male* sebesar 21,2%.



Gambar 5 *Race/Ethnicity* berdasarkan Nilai

Gambar 5: dapat kita lihat bahwa grup dengan *score math* paling tinggi dimiliki oleh *Group E*, kemudian dengan *score reading* tertinggi dimilki oleh *Group D* dan *Group E*, dan terakhir pada *score writing* tertinggi dimiliki oleh *Group D* dan *Group E*.

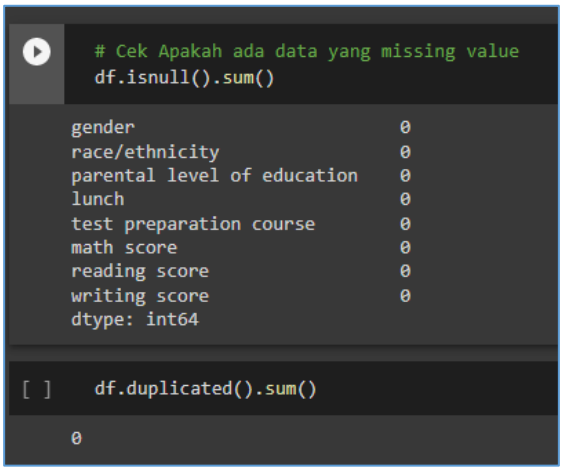


Gambar 6 Penyebaran nilai berdasarkan *gender*

Gambar 6 yaitu variabel *male* dan *female* pada *math score* saling berdekatan(tidak beraturan), variabel *male* sedikit dominan. Kemudian variabel *male* dan *female* pada *reading score* saling menindih sehingga variabel *male* menjadi sangat dominan. Dan variabel *male* dan *female* pada *writing score* saling berdekatan(beraturan), 2 variabel seimbang.

3.2.3 *Data Preprocessing*

Data preprocessing adalah teknik awal *data mining* untuk mengubah *raw data*(data mentah) menjadi format dan informasi yang lebih efisien dan bermanfaat. Format pada *raw data* yang diambil dari berbagai macam sumber seringkali mengalami error, *missing value*, dan tidak konsisten. Sehingga, perlu dilakukan pembenahan format agar hasil data mining tepat dan akurat. Pada kasus *student performance* tidak memiliki data yang kosong, *error*, data terduplikat, dan lain sebagainya.



Gambar 7 *Preprocessing*

Feature Engineering adalah bagaimana kita menggunakan pengetahuan kita dalam memilih *features* atau membuat *features* baru agar model *machine learning* dapat bekerja lebih akurat dalam memecahkan masalah.

math score	reading score	writing score
72	72	74
69	90	88
90	95	93
47	57	44
76	78	75

Gambar 8 Sebelum *feature engineering*

Pada kasus proyek diatas penulis menambah kolom seperti *total score*, *average score*, dan yang terakhir adalah keterangan yang berdasarkan *average score* yang akan menentukan lulus atau tidaknya, berikut dibawah adalah gambar sesudah melakukan *feature engineering*.

math score	reading score	writing score	total_score	avg_nilai	keterangan
72	72	74	218	72.666667	lulus
69	90	88	247	82.333333	lulus
90	95	93	278	92.666667	lulus
47	57	44	148	49.333333	tidak lulus
76	78	75	229	76.333333	lulus

Gambar 9 *Feature Engineering*

Label Encoder digunakan untuk mengonversi data kategorikal, atau data teks, menjadi angka, yang dapat lebih dipahami oleh model. Berikut adalah hasil setelah di *encoder*.

Gender	Hasil Encoder
Female	0
Male	1

Keterangan	Hasil Encoder
Lulus	0
Tidak Lulus	1

Race/Ethnicity	Hasil Encoder
Group A	0
Group B	1
Group C	2
Group D	3
Group E	4

Test Preparation course	Hasil Encoder
Completed	0
None	1

Lunch	Hasil Encoder
Free/Reduced	0
Standard	1

Parental level of education	Hasil Encoder
Associate's Degree	0
Bachelor's Degree	1
High School	2
Master's Degree	3
Some College	4
Some High School	5

Gambar 10 *Label Encoder*

Train-Test Split adalah membagi dataset menjadi *train set* dan *test set*, atau dengan kata lain, data yang digunakan untuk proses *training* dan *testing* merupakan kumpulan data yang berbeda. Dalam membagi *test size* pada umumnya sebesar 20% - 30%. *Feature Scaling* adalah suatu cara untuk membuat *numerical data* pada dataset memiliki rentang nilai (*scale*) yang sama. Tidak ada lagi satu variabel data yang mendominasi variabel data lainnya. Adapun alasan penulis memilih *standarization* dikarenakan cocok pada data yang sifatnya terdistribusi normal.

Standardisation	Normalisation
$x_{stand} = \frac{x - \text{mean}(x)}{\text{standard deviation } (x)}$	$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)}$

Gambar 11 *Feature Scaling*

3.2.4. Evaluation and Recommendation Model

Machine learning model adalah algoritma *machine learning* yang sebelumnya telah dilakukan proses pelatihan/*training* dengan data latih tertentu sehingga dia siap digunakan untuk melakukan prediksi terhadap data baru. Adapun algoritma yang diterapkan ialah *KNearest Neighbors* (KNN) adalah algoritma yang berfungsi untuk melakukan klasifikasi suatu data berdasarkan data pembelajaran (*train data sets*), yang diambil dari *k* tetangga terdekatnya (*nearest neighbors*). Dengan *k* merupakan banyaknya tetangga terdekat. Lalu yang kedua adalah algoritma *Support Vector Machine* (SVM) digunakan untuk mencari *hyperplane* terbaik dengan memaksimalkan jarak antar kelas. Dan yang terakhir adalah algoritma *Multilayer Perceptron* (MLP) adalah salah satu permodelan dalam teknologi jaringan saraf tiruan (JST) dengan karakteristik memiliki nilai bobot yang lebih baik dari pada pemodelan yang lain, sehingga menghasilkan klasifikasi yang lebih akurat. Berikut adalah hasil setelah melakukan pemodelan dengan beberapa algoritma.

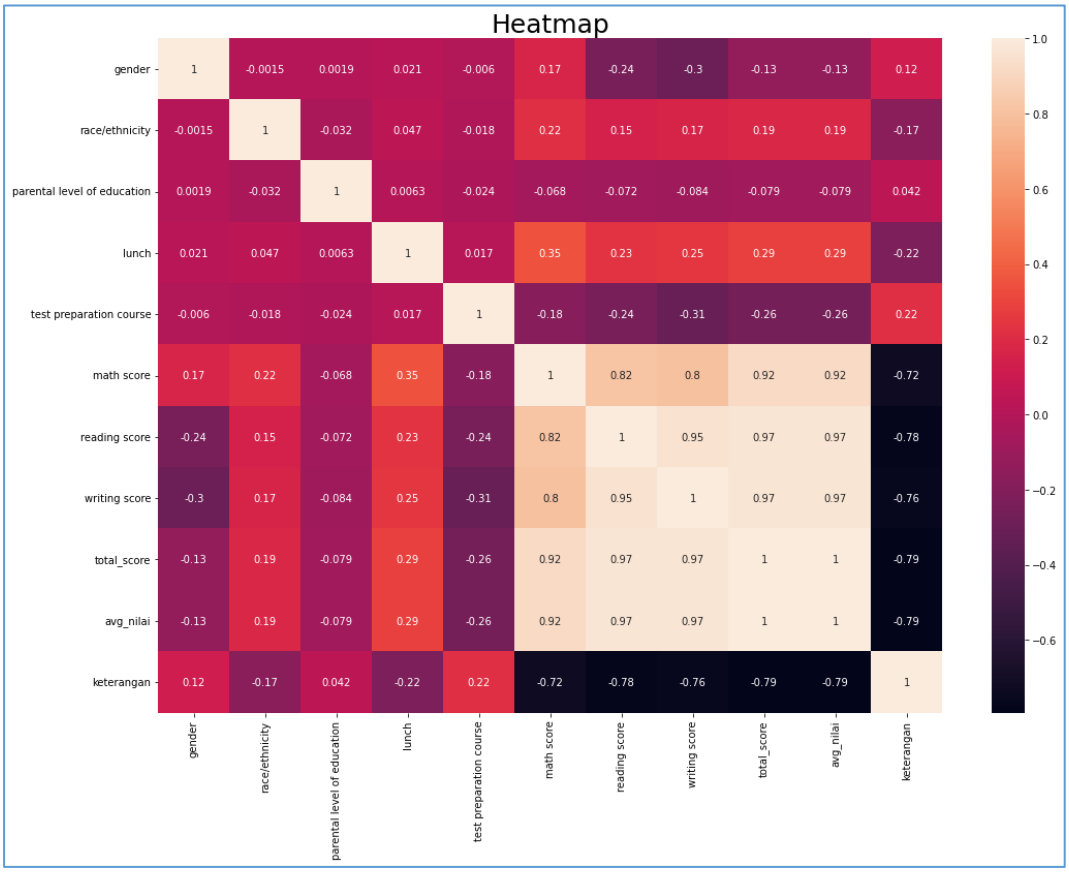
Train-Test Spli	**KNN**	**SVM**	**MLP**
TRAINING	*Accuracy* : 97.38 % *Precision* : 97.43 % *Recall* : 97.66 %	*Accuracy* : 98.62 % *Precision* : 97.71 % *Recall* : 99.77 %	*Accuracy* : 99.88 % *Precision* : 99.77 % *Recall* : 100.0 %
TESTING	*Accuracy* : 92 % *Precision* : 92.66 % *Recall* : 92.66 %	*Accuracy* : 98.0 % *Precision* : 98.17 % *Recall* : 98.17 %	*Accuracy* : 99.5 % *Precision* : 100.0 % *Recall* : 99.08 %

Kemudian untuk mengotomatisasi/memaksimalkan kinerja model maka penulis menggunakan *machine learning pipeline*, dan berikut adalah hasilnya.

KNN	**SVM**	**MLP**
Accuracy : 92. % *Precision* : 92.66 % *Recall* : 92.66 %	*Accuracy* : 98.0 % *Precision* : 98.17 % *Recall* : 98.17 %	*Accuracy* : 99 % *Precision* : 99 % *Recall* : 99 %

III.3. Rekomendasi Model *Machine Learning*

Berdasarkan hasil penerapan pemodelan *machine learning* yang sudah penulis lakukan maka merekomendasikan model *Support vektor machine* (SVM), karena perbandingan nilai *accuracy* untuk evaluasi model dan pemodelan *pipeline* tidak berubah atau tidak mengalami perubahan yang sangat jauh artinya model tersebut bisa diterapkan.



Gambar 12 Korelasi antar kolom

Kemudian berdasarkan *heatmap* diatas, maka korelasi *lunch* dan *math score* memiliki nilai paling tinggi dibandingkan variabel lain, dalam arti nilai matematika juga dipengaruhi oleh makan siang.

Bab IV Penutup

IV.1. Kesimpulan

Kesimpulan yang dapat diambil dari pelaksanaan MSIB di PT Digitalisasi Pemuda Indonesia atau Digital Skola antara lain:

1. Mampu beradaptasi dengan baik sehingga terjalin hubungan antar peserta dan tim Digital Skola.

2. Tim Digital Skola mengadakan 2 proyek utama yaitu prediksi klasifikasi dan prediksi *time series forecasting*.

3. Hasil yang telah dilakukan oleh penulis pada studi kasus *education student performance* bisa diterapkan bagi siswa yang ingin memprediksi apakah nilai berpengaruh pada variabel lain, misal suku/ras, jenis kelamin, pendidikaan orang tua, dan lain sebagainya.

IV.2. Saran

Adapun saran dari pelaksanaan MSIB di PT Digitalisasi Pemuda Indonesia atau Digital Skola antara lain:

1. Untuk pengerjaan proyek akhir penulis berharap diadakannya proyek klasifikasi berupa text ataupun gambar.

2. Perlu adanya perbaikan pemodelan *machine learning* pada studi kasus diatas, misalnya menambah *feature* baru agar model bisa banyak belajar pada data baru. Kemudian bisa menambah *hyperparameter tuning* agar kinerja model lebih maksimal.

3. Tentunya penulis sudah menyadari jika dalam penyusunan laporan ini masih banyak ada kesalahan serta jauh dari kata sempurna. Adapun nantinya penulis akan segera melakukan perbaikan susunan laporan ini dengan menggunakan pedoman dari beberapa sumber dan kritik yang bisa membangun dari para pembaca.

References

[1] C. Indonesia, *Kampus Merdeka ala Mas Nadiem Makarim di Era Revolusi Industri 4.0,* 2020.

[2] C. Darujati and A. B. Gumelar, *Pemanfaatan Teknik Supervised Untuk Klasifikasi Teks Bahasa Indonesia,* pp. 1-5, 2012.

[3] "Direktorat Jenderal Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan," *Buku Panduan Merdeka Belajar - Kampus Merdeka,* 2020.

[4] "Kementerian Pendidikan dan Kebudayaan," *Merdeka Belajar: Kampus Merdeka,* 2020.

[5] G. Kosala, A. Harjoko and S. Hartati, "Proceedings of the International Conference on Video and Image Processing," *License Plate Detection Based on Convolutional Neural Network: SupportVector Machine (CNN-SVM),* pp. 1-5, 2017.

[6] Kusrini and E. T. Luthfi, *Algoritma Data Mining,* 2009.

[7] H. Lubaba, *Merdeka dalam Berpikir,* 2020.

[8] Y. Makdori, *Kemdikbud Rilis Konsep Kampus Merdeka untuk Perguruan Tinggi,* 2020.

[9] Noertjahyana, Agustinus and Yulia, "Studi Analisa Pelatihan Jaringan Syaraf Tiruan dengan dan tanpa Algoritma Genetika," *Jurnal Informatika,* vol. 3, pp. 13-18, 2002.

[10] P. A. Octaviani, Y. Wilandari and D. Ispriyanti, "Penerapan Metode Klasifikasi Support Vector Machine (SVM) Pada Data Akreditasi Sekolah Dasar (SD) DiKabupaten Magelang," *Jurnal Gaussian,* pp. 811-820, 2014.

[11] K. Teknomo, "K-Nearest Neighbours Tutorial, K-Tetangga Terdekat Tutorial," 2010.

[12] R. A. Wattimena, "Kemerdekaan Pikiran," 2012.

Lampiran A. TOR Digital Skola

A.1. *Quiz*

B.1. *Homework*

C.1. *Learning Progress Review*

D.1. *Personal Branding*

E.1. *Dataset & Modelling*

Capaian Pembelajaran Lulusan

Topik	Tingkat Kompetensi	Detail Pembelajaran	Durasi Total Jam (Belajar/ Bimbingan & Tugas)	Sumber Daya Pembelajaran	Cara Penilaian
Analytic	Memahami materi yang terkait dengan *Analytics*, yang meliputi metodologi *data science*, Numpy, *dataframe*, statistik, visualisasi data, dan *business intelligence*.	*1. Introduction to data Science* *2. Data Science Methodology* *3. Introduction to Numpy* *4.Introduction and basic dataframe (Pandas)* *5. Intermediate dataframe* *6. Advance dataframe* *7. Business intelligence*	60	1. *Interactive live-online learning* (dengan *Expert* maupun Pembimbing) 2. Silabus pembelajaran (PPT) 3. Video *recording* (pengulangan materi)	Peserta diberikan tugas berupa: 1. Kuis 2. PR (*essay*/koding) 3. Merangkum materi
Data & Database	Memahami segala hal tentang data dan *database*, termasuk di dalamnya *Structured Query Language* (SQL), *version control, Application Programming Interface (API), and database programming*	*1. Introduction to data and database* *2. Basic SQL* *3. Intermediate SQL* *4. Advance SQL* *5. Versioning/version control* *6. Data & Database* *7. API*	73	1. *Interactive live-online learning* (dengan *Expert* maupun Pembimbing) 2. Silabus pembelajaran (PPT) 3. Video *recording* (pengulangan materi)	Peserta diberikan tugas berupa: 1. Kuis 2. PR (*essay*/koding) 3. Merangkum materi

Python Programming	Menguasai bahasa pemrograman Python termasuk *libraries* yang digunakan	*1. Introduction to python and programming* *2. Basic programming I: conditions* *3. Basic programming II: Iteration* *4. Basic Programming III: Array and Other data types* 5. *Basic Programming IV: functions*	52	1. *Interactive live-online learning* (dengan *Expert* maupun Pembimbing) 2. Silabus pembelajaran (PPT) 3. Video *recording* (pengulangan materi)	Peserta diberikan tugas berupa: 1. Kuis 2. PR (*essay*/koding) 3. Merangkum materi
Statistics	Memahami dasar- dasar statistika yang digunakan pada domain *data science*	*1. Basic statistic* *2. Intermediate statistic* 3. *Advanced statistic*	31	1. *Interactive live-online learning* (dengan *Expert* maupun Pembimbing) 2. Silabus pembelajaran (PPT) 3. Video *recording* (pengulangan materi)	Peserta diberikan tugas berupa: 1. Kuis 2. PR (*essay*/koding) 3. Merangkum materi
Data Visualization	Mampu membuat visualisasi data secara efektif	*1. Introduction to data visualization* *2. Intermediate visualization* 3. *Data visualization exercise*	42	1. *Interactive live-online learning* (dengan *Expert* maupun Pembimbing) 2. Silabus pembelajaran (PPT) 3. Video *recording* (pengulangan materi)	Peserta diberikan tugas berupa: 1. Kuis 2. PR (*essay*/koding) 3. Merangkum materi

Machine Learning	Mampu membuat model *machine learning* sesuai kebutuhan	*1. Introduction to data mining* *2. Introduction to machine learning* *3. Data preprocessing for ML (*python*)* *4. Advanced data preprocessing for ML (python)* *5. Classification I* *6. Classification II* *7. Regression* *8. Unsupervised learning* *9. Evaluation metrics & model selection* *10. Advanced machine learning topics*	104	1. *Interactive live-online learning* (dengan *Expert* maupun Pembimbing) 2. Silabus pembelajaran (PPT) 3. Video *recording* (pengulangan materi)	Peserta diberikan tugas berupa: 1. Kuis 2. PR (*essay*/koding) 3. Merangkum materi
Artificial Intelligence	Memahami cara penggunaan teknologi AI dan mampu mengimplementasikan nya pada bisnis	*1. Introduction to Artificial Intelligence* *2. Advanced data preprocessing for text* *3. Advanced data preprocessing for image* *4. neural network* *5. Classification for text dataset* *6. Classification for image dataset* *7. Time Series forecasting* *8. Unsupervised Learning II* *9. Semi-supervised learning* *10 Association rules 11.Outlier detection method* *12. Recommender system*	124	1. *Interactive live-online learning* (dengan *Expert* maupun Pembimbing) 2. Silabus pembelajaran (PPT) 3. Video *recording* (pengulangan materi)	Peserta diberikan tugas berupa: 1. Kuis 2. PR (*essay*/koding) 3. Merangkum materi

Soft Skill	*Communication & Presentation Skill* : Mampu berkomunikasi secara efektif, baik komunikasi individual maupun komunikasi kepada publik	Pemahaman mengenai: *1. Story Telling* 2. *Effective presentation*	36	1. *Interactive live-online learning* (dengan *Expert* maupun Pembimbing) 2. Silabus pembelajaran (PPT) 3. Video *recording* (pengulangan materi)	Peserta diberikan tugas berupa: 1. Kuis 2. Merangkum Materi
	Analytical & Critical Thinking: Mampu melihat masalah dari sisi user serta memberikan solusi yang tepat	Pemahaman mengenai: *1. Design Thinking* *2. Design thinking steps* *3. Design thinking tools* *4. Assume a beginner's mindset* *5. Ask the 5 whys* 6. *Empthy map*		1. *Interactive live-online learning* (dengan *Expert* maupun Pembimbing) 2. Silabus pembelajaran (PPT) 3. Video *recording* (pengulangan materi)	Peserta diberikan tugas berupa: 1. Kuis 2. Merangkum Materi
	Career Coaching & Mentoring: Mampu membuat CV yang efektif dan mampu membuat kesan yang baik pada saat di *interview*	Pemahaman mengenai: *1. How to make a powerful CV* *2. How to make a good impression in an interview* 3. *How to express your desired benefits in an interview*		1. *Interactive live-online learning* (dengan *Expert* maupun Pembimbing) 2. Silabus pembelajaran (PPT) 3. Video *recording* (pengulangan materi)	Peserta diberikan tugas berupa: 1. Kuis 2. Merangkum Materi

Dataset & Modelling Project	Mampu mengerjakann *Project dataset* yang diberikan serta membuat *modeling*-nya	Pemahaman mengenai: *1. How to understand the dataset given* *2. How to identify which activities should be done* *3. using exploratory and visualization* *4. How to do data pre-processing* *5. How to develop model which relevant with the problem* 6. *How to evaluate the model*	97	1. *Interactive live-online learning* (dengan *Expert* maupun Pembimbing) 2. Silabus pembelajaran (PPT) 3. Video *recording* (pengulangan materi)	Peserta diberikan tugas berupa: 1. Mengerjakan *dataset* dan membuat modeling hasil modeling yangtelah dibuat 2. Menjawab pertanyaanpenilai (juri)

Lampiran B. Log Activity

Minggu/Tgl	Kegiatan	Hasil
24 Agustus 2021	*Class Introduction*	Selesai
26 Agustus 2021	*Introduction to Data Science*	Selesai
28 Agustus 2021	*Data Science Methodology*.	Selesai
31 Agustus 2021	*Introduction to Data and Database*	Selesai
2 September 2021	*Basic SQL*	Selesai
4 September 2021	*Intermediate SQL*	Selesai
7 September 2021	*Advanced SQL*	Selesai
9 September 2021	*Versioning/Version Control*	Selesai
11 September 2021	*Introduction to Python and Programming.*	Selesai
14 September 2021	*Basic Programming I: Conditions,*	Selesai
16 September 2021	*Basic Programming II: Iteration*	Selesai

18 September 2021	*Basic Programming III: Array and Other Data Types.*	Selesai
21 September 2021	*Basic Programming IV: Functions*	Selesai
23 September 2021	*Database Programming*	Selesai
25 September 2021	*Introduction to Numpy*	Selesai
28 September 2021	*Introduction and Basic Dataframe (Pandas)*	Selesai
30 September 2021	*Kaggle Project Stage 1*	Selesai
2 Oktober 2021	*Analytical & Critical Thinking*	Selesai
5 Oktober 2021	*Intermediate Dataframe I*	Selesai
7 Oktober 2021	*Advanced Dataframe*	Selesai
9 Oktober 2021	*API*	Selesai
12 Oktober 2021	*Basic Statistics,*	Selesai
14 Oktober 2021	*Intermediate Statistics*	Selesai

16 Oktober 2021	*Advanced Statistics*	Selesai
19 Oktober 2021	*Introduction to Data Visualization,*	Selesai
21 Oktober 2021	*Intermediate Visualization*	Selesai
23 Oktober 2021	*Data Visualization Exercises (Advance).*	Selesai
26 Oktober 2021	*Introduction to Data Mining*	Selesai
28 Oktober 2021	*Introduction to Machine Learning*	Selesai
30 Oktober 2021	*Data Preprocessing for ML (python).*	Selesai
2 November 2021	*Advanced Data Preprocessing for ML (python)*	Selesai
4 November 2021	*Classification I*	Selesai
6 November 2021	*Classification II*	Selesai
9 November 2021	*Regression*	Selesai
11 November 2021	*Unsupervised Learning*	Selesai
13 November 2021	*Communication & Presentation Skill*	Selesai

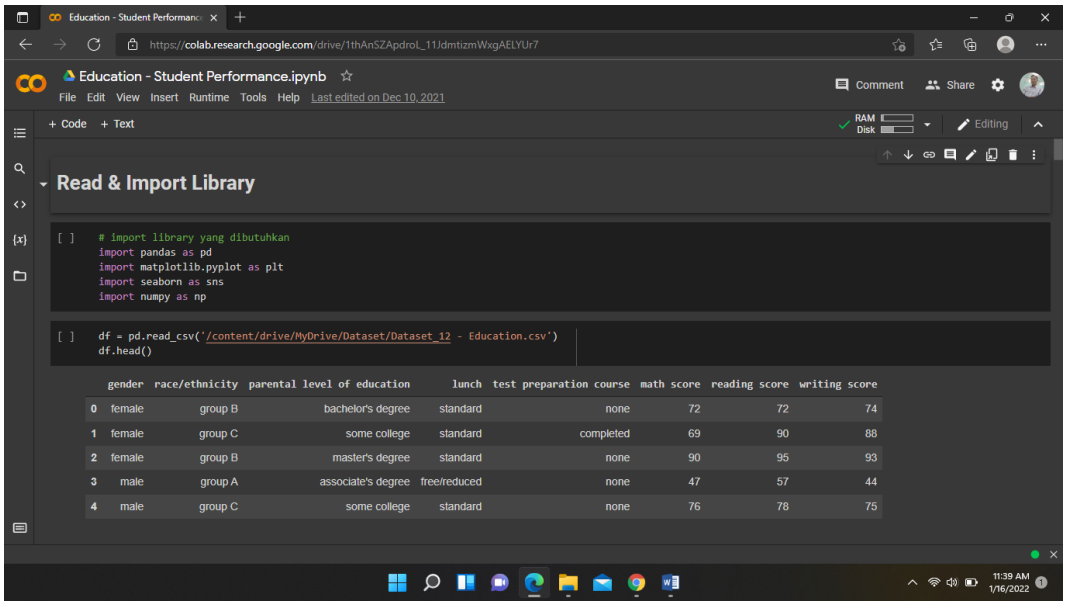
16 November 2021	*Evaluation Metrics and Model Selection*	Selesai
18 November 2021	*Advanced ML Topics*	Selesai
20 November 2021	*Business Intelligence*	Selesai
23 November 2021	*Mentor Experience Sharing*	Selesai
25 November 2021	*HR Practicioner Sharing*	Selesai
27 November 2021	*Kaggle Project Presentation*	Selesai
29 November 2021	*Kaggle Project Presentation*	Selesai
2 Desember 2021	*Kaggle Project Presentation*	Selesai
5 Desember 2021	*Introduction to Artificial Intelligence.*	Selesai
7 Desember 2021	*Advanced Data Preprocessing for Text,*	Selesai
9 Desember 2021	*Advanced Data Preprocessing for Image*	Selesai

11 Desember 2021	*Neural Network*	Selesai
14 Desember 2021	*Classification for Text Dataset*	Selesai
16 Desember 2021	*Classification for Image Dataset,*	Selesai
18 Desember 2021	*Time series Forecasting*	Selesai
21 Desember 2021	*Unsupervised Learning II*	Selesai
23 Desember 2021	*Semi-supervised Learning*	Selesai
28 Desember 2021	*Association Rules*	Selesai
30 Desember 2021	*Outlier Detection Method*	Selesai
4 Januari 2022	*Recommender System*	Selesai
20 Januari 2022	*Kaggle Project Presentation*	Selesai
22 Januari 2022	*Kaggle Project Presentation*	Selesai
25 Januari 2022	*Kaggle Project Presentation*	Selesai
26 Januari 2022	*Finalisasi Draft Laporan*	Selesai
27 Januari 2022	*Finalisasi Draft Laporan*	Selesai

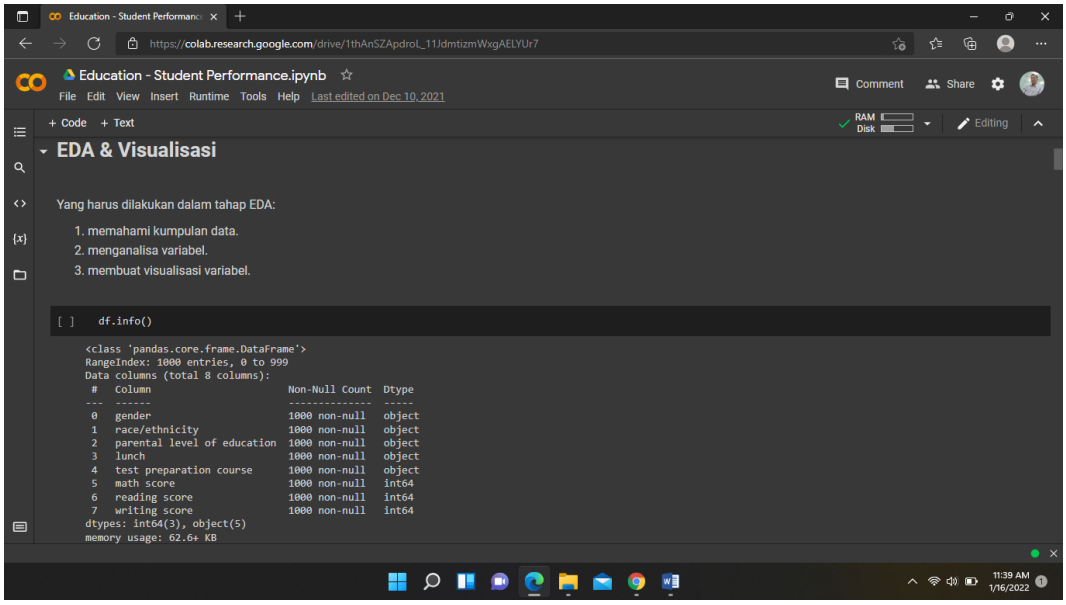
Lampiran C. Dokumen Teknik

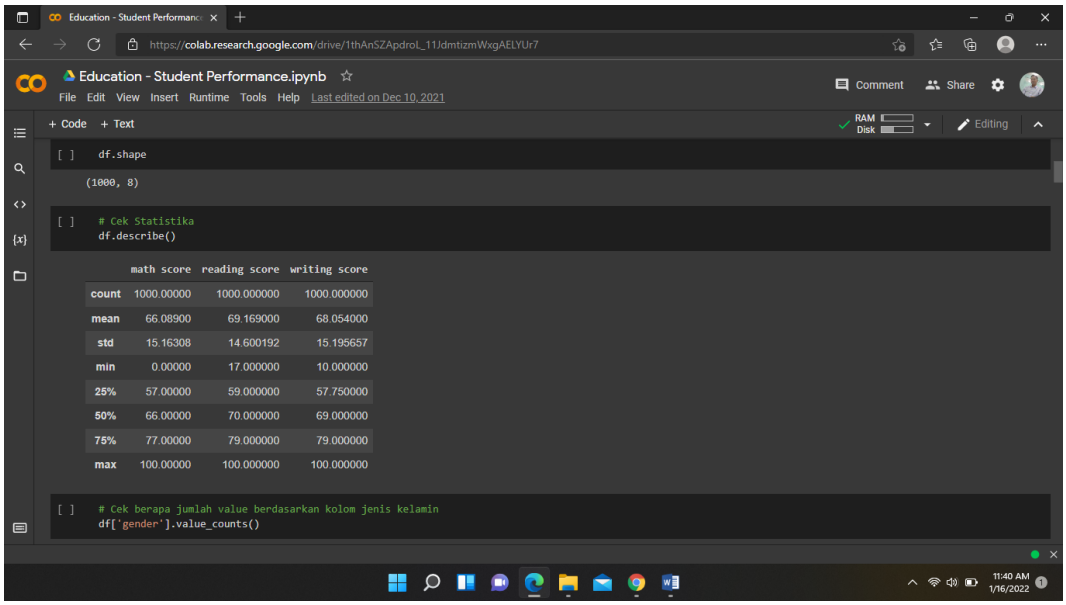
Berikut ini adalah *screenshot* hasil *google colab* dalam pengerjaan *project* program MSIB Digital Skola:

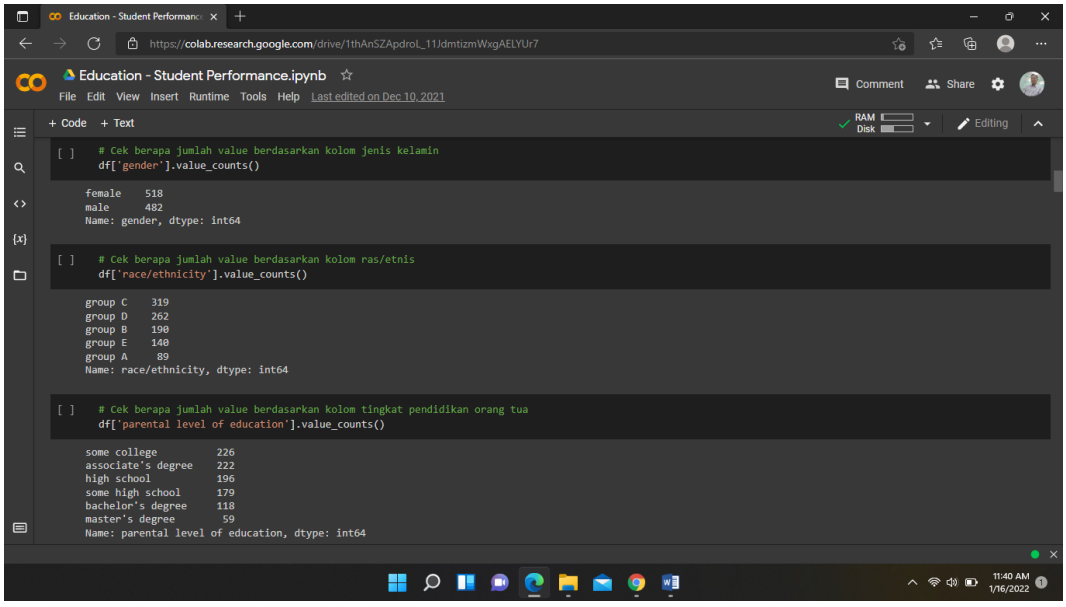
C.1. Import libraries and Read Dataset

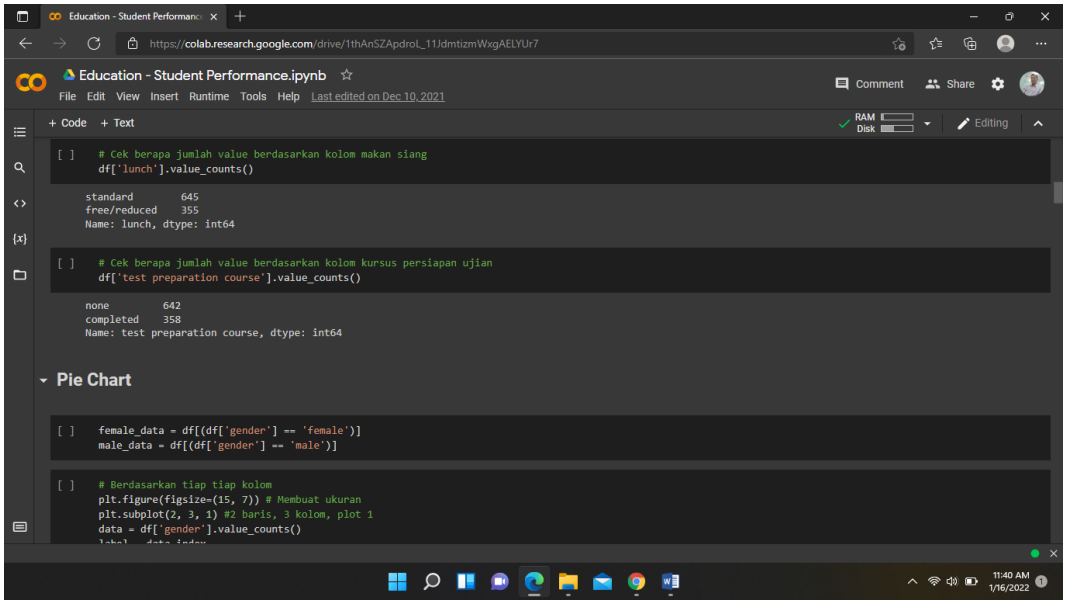


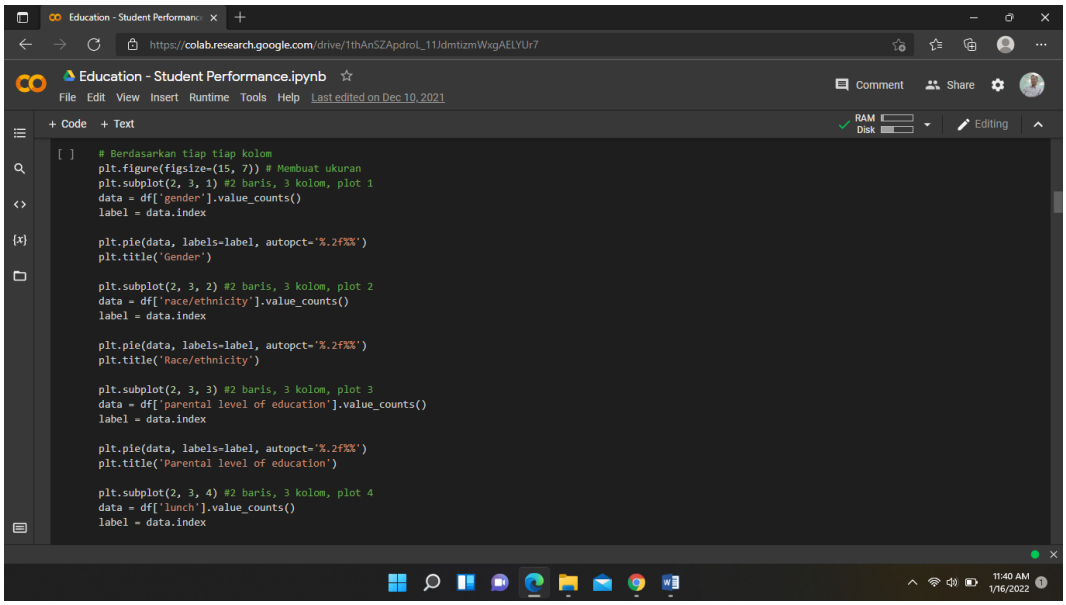
I.2. Exploratory Data Analysis & Visualization

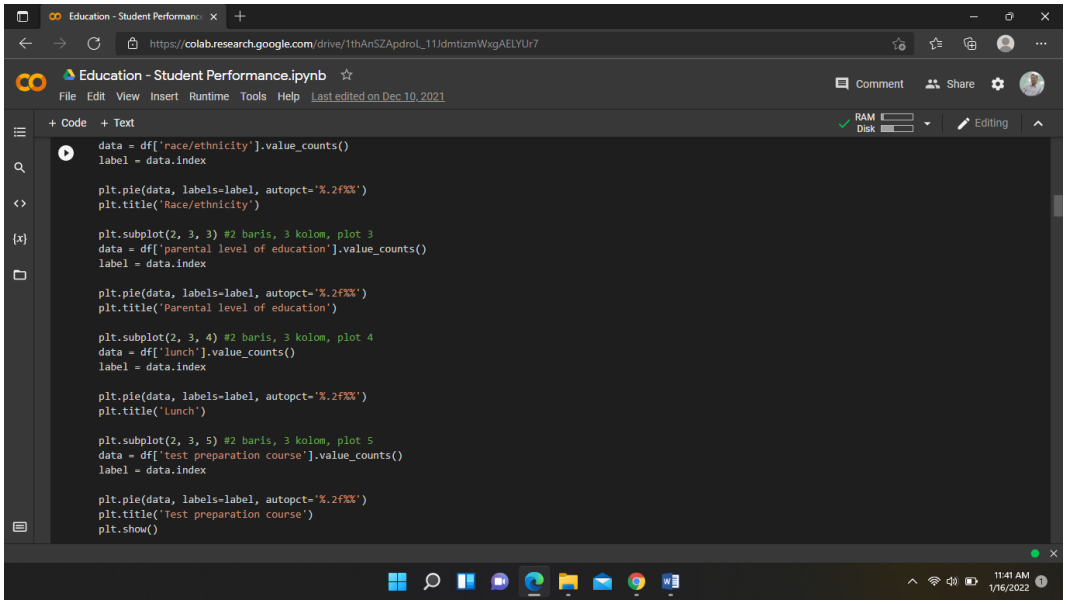


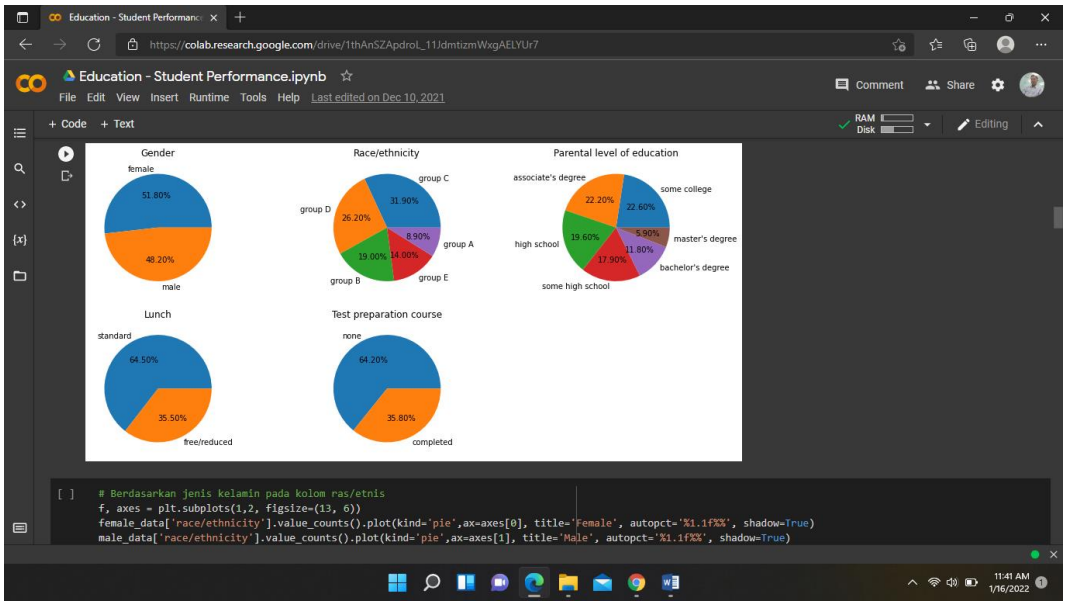

Education - Student Performanc
https://colab.research.google.com/drive/1thAnSZApdroL_11JdmtizmWxgAELYUr7
Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021
Comment
Share
+ Code + Text
RAM
Disk
Editing
[ ] df.shape
(1000, 8)
[ ] # Cek Statistika
df.describe()
math score reading score writing score
count 1000.00000 1000.000000 1000.000000
mean 66.08900 69.169000 68.054000
std 15.16308 14.600192 15.195657
min 0.00000 17.000000 10.000000
25% 57.00000 59.000000 57.750000
50% 66.00000 70.000000 69.000000
75% 77.00000 79.000000 79.000000
max 100.00000 100.000000 100.000000
[ ] # Cek berapa jumlah value berdasarkan kolom jenis kelamin
df['gender'].value_counts()
11:40 AM
1/16/2022

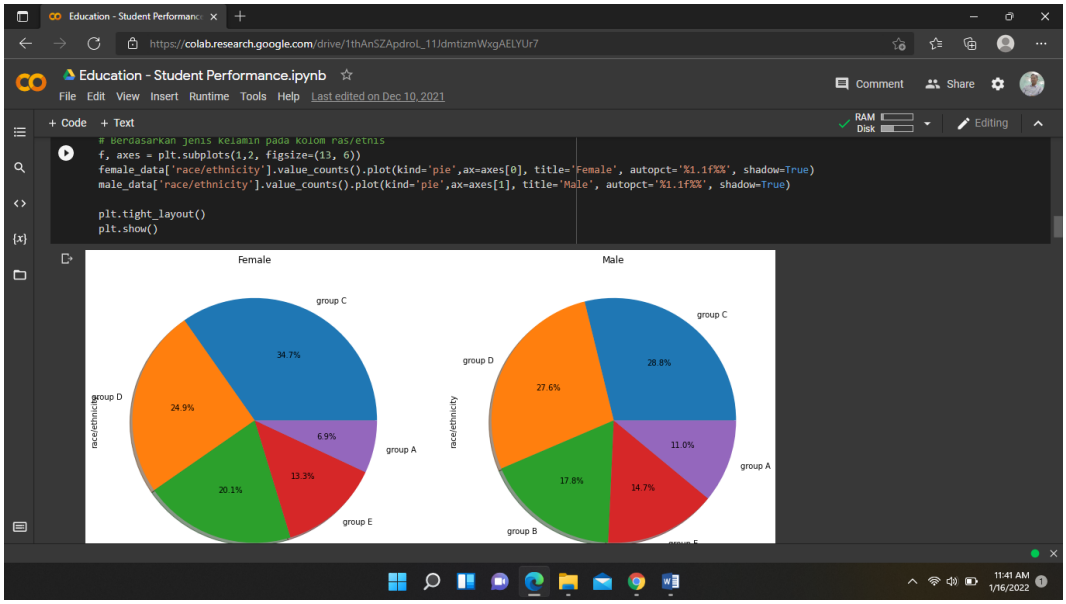

Education - Student Performanc
https://colab.research.google.com/drive/1thAnSZApdroL_11JdmtizmWxgAELYUr7
Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021
Comment
Share
+ Code + Text
RAM
Disk
Editing
[ ] # Cek berapa jumlah value berdasarkan kolom jenis kelamin
df['gender'].value_counts()
female 518
male 482
Name: gender, dtype: int64
[ ] # Cek berapa jumlah value berdasarkan kolom ras/etnis
df['race/ethnicity'].value_counts()
group C 319
group D 262
group B 190
group E 140
group A 89
Name: race/ethnicity, dtype: int64
[ ] # Cek berapa jumlah value berdasarkan kolom tingkat pendidikan orang tua
df['parental level of education'].value_counts()
some college 226
associate's degree 222
high school 196
some high school 179
bachelor's degree 118
master's degree 59
Name: parental level of education, dtype: int64
11:40 AM
1/16/2022

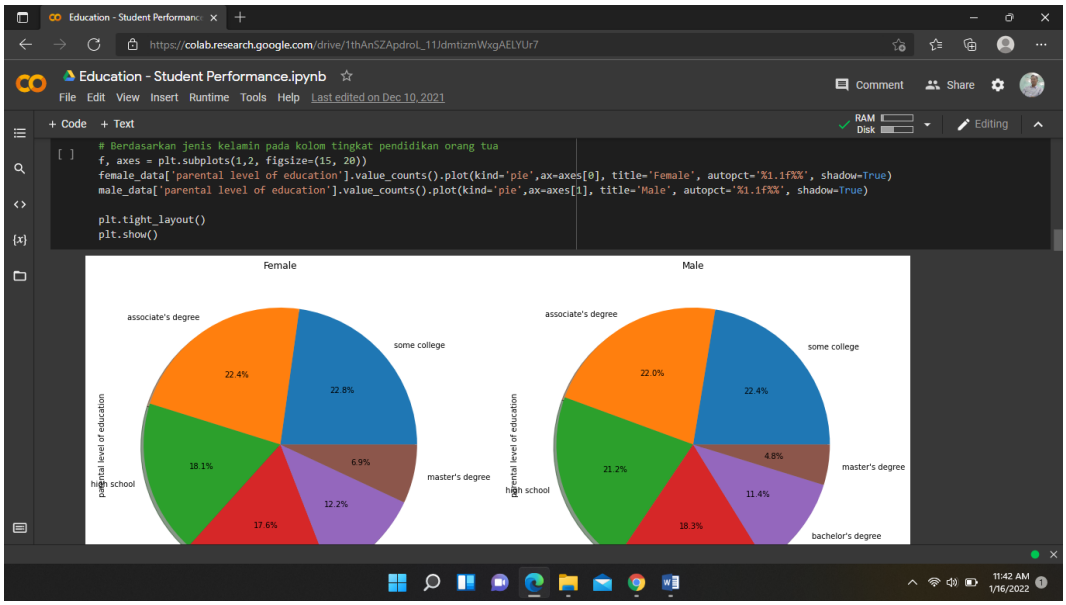

Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021
Comment
Share
+ Code + Text
RAM
Disk
Editing
[ ] # Cek berapa jumlah value berdasarkan kolom makan siang
df['lunch'].value_counts()
standard 645
free/reduced 355
Name: lunch, dtype: int64
[ ] # Cek berapa jumlah value berdasarkan kolom kursus persiapan ujian
df['test preparation course'].value_counts()
none 642
completed 358
Name: test preparation course, dtype: int64
Pie Chart
[ ] female_data = df[(df['gender'] == 'female')]
male_data = df[(df['gender'] == 'male')]
[ ] # Berdasarkan tiap tiap kolom
plt.figure(figsize=(15, 7)) # Membuat ukuran
plt.subplot(2, 3, 1) #2 baris, 3 kolom, plot 1
data = df['gender'].value_counts()
11:40 AM
1/16/2022

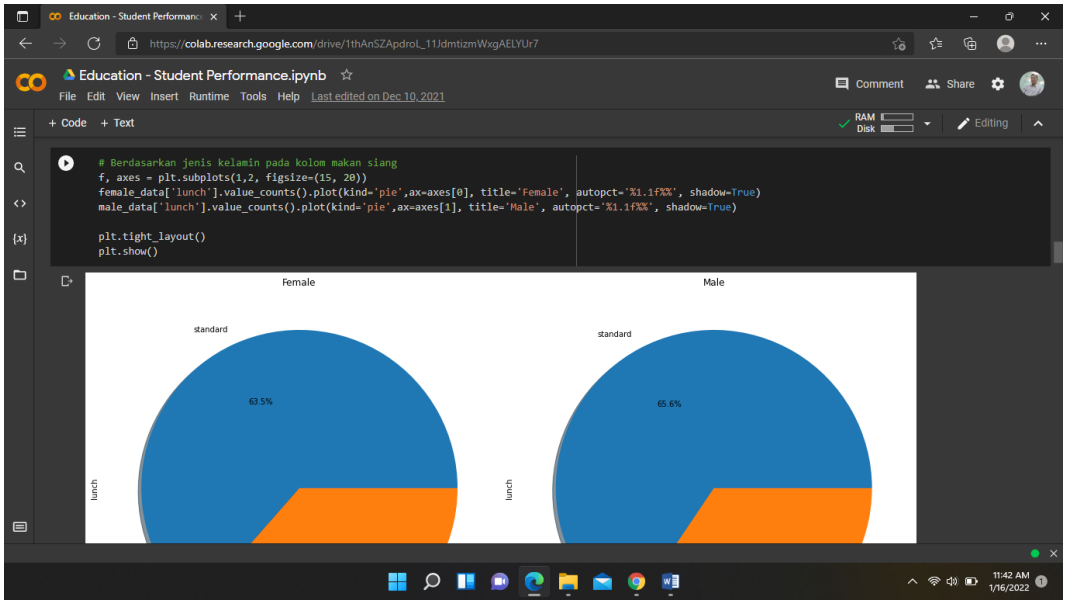

Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021
Comment
Share
+ Code + Text
RAM
Disk
Editing
[ ] # Berdasarkan tiap tiap kolom
plt.figure(figsize=(15, 7)) # Membuat ukuran
plt.subplot(2, 3, 1) #2 baris, 3 kolom, plot 1
data = df['gender'].value_counts()
label = data.index
plt.pie(data, labels=label, autopct='%.2f%%')
plt.title('Gender')
plt.subplot(2, 3, 2) #2 baris, 3 kolom, plot 2
data = df['race/ethnicity'].value_counts()
label = data.index
plt.pie(data, labels=label, autopct='%.2f%%')
plt.title('Race/ethnicity')
plt.subplot(2, 3, 3) #2 baris, 3 kolom, plot 3
data = df['parental level of education'].value_counts()
label = data.index
plt.pie(data, labels=label, autopct='%.2f%%')
plt.title('Parental level of education')
plt.subplot(2, 3, 4) #2 baris, 3 kolom, plot 4
data = df['lunch'].value_counts()
label = data.index
11:40 AM
1/16/2022

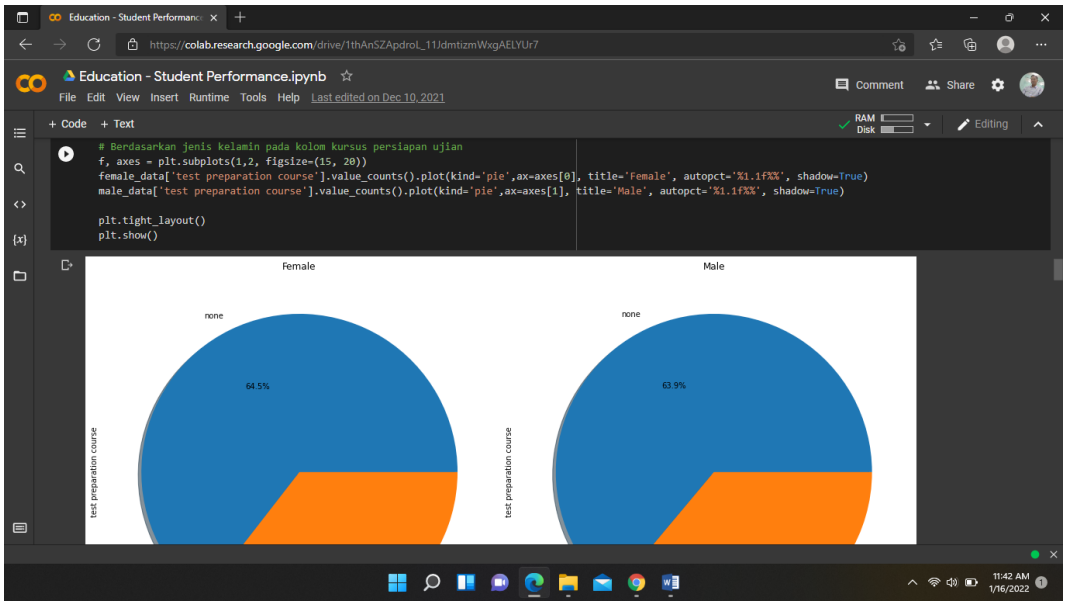

Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021
Comment Share
+ Code + Text
RAM Disk
Editing
data = df['race/ethnicity'].value_counts()
label = data.index
plt.pie(data, labels=label, autopct='%.2f%%')
plt.title('Race/ethnicity')
plt.subplot(2, 3, 3) #2 baris, 3 kolom, plot 3
data = df['parental level of education'].value_counts()
label = data.index
plt.pie(data, labels=label, autopct='%.2f%%')
plt.title('Parental level of education')
plt.subplot(2, 3, 4) #2 baris, 3 kolom, plot 4
data = df['lunch'].value_counts()
label = data.index
plt.pie(data, labels=label, autopct='%.2f%%')
plt.title('Lunch')
plt.subplot(2, 3, 5) #2 baris, 3 kolom, plot 5
data = df['test preparation course'].value_counts()
label = data.index
plt.pie(data, labels=label, autopct='%.2f%%')
plt.title('Test preparation course')
plt.show()

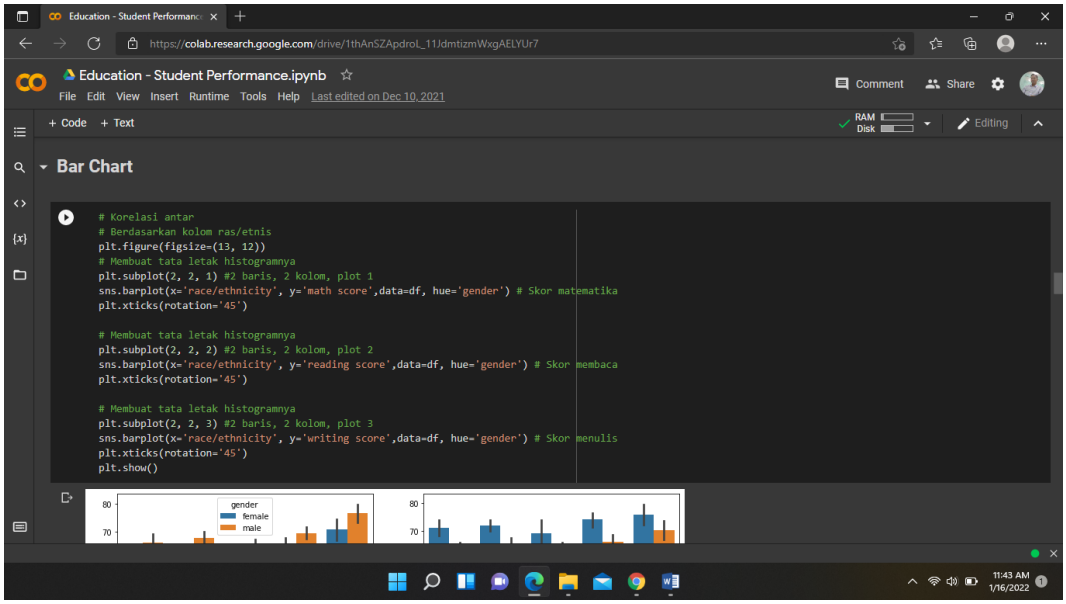

Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021
Comment Share
+ Code + Text
RAM Disk
Editing
Gender
female
51.80%
48.20%
male
Race/ethnicity
group C
31.90%
group D
26.20%
8.90%
group A
19.00%
14.00%
group B
group E
Parental level of education
associate's degree
some college
22.20%
22.60%
5.90%
high school
19.60%
master's degree
11.80%
17.90%
bachelor's degree
some high school
Lunch
standard
64.50%
35.50%
free/reduced
Test preparation course
none
64.20%
35.80%
completed
[ ] # Berdasarkan jenis kelamin pada kolom ras/etnis
f, axes = plt.subplots(1,2, figsize=(13, 6))
female_data['race/ethnicity'].value_counts().plot(kind='pie',ax=axes[0], title='Female', autopct='%1.1f%%', shadow=True)
male_data['race/ethnicity'].value_counts().plot(kind='pie',ax=axes[1], title='Male', autopct='%1.1f%%', shadow=True)

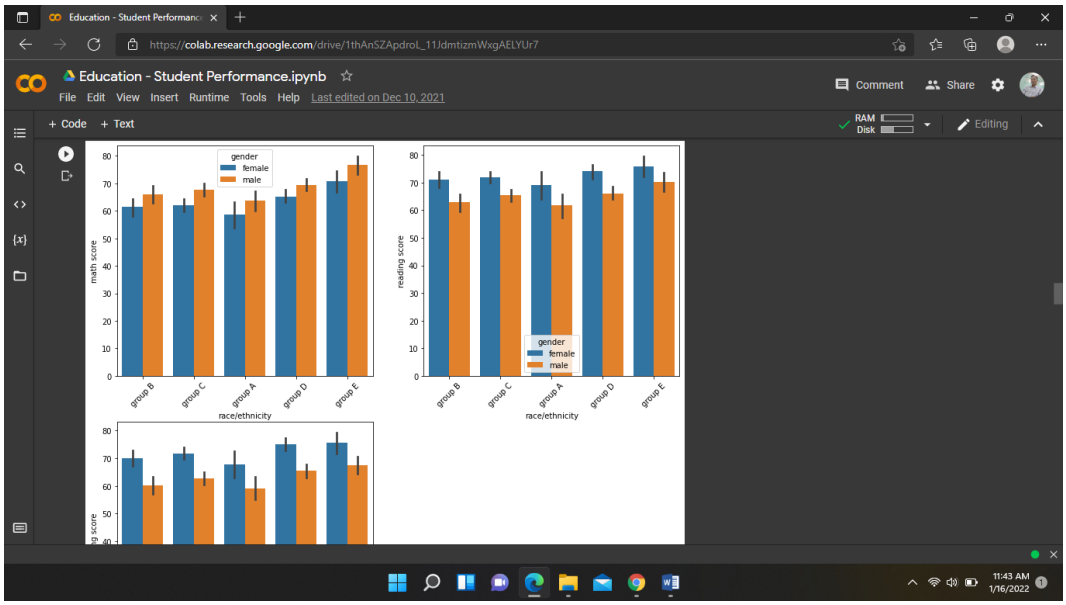

Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021
Comment Share
+ Code + Text
RAM Disk
Editing
Berdasarkan jenis kelamin pada kolom ras/etnis
f, axes = plt.subplots(1,2, figsize=(13, 6))
female_data['race/ethnicity'].value_counts().plot(kind='pie',ax=axes[0], title='Female', autopct='%1.1f%%', shadow=True)
male_data['race/ethnicity'].value_counts().plot(kind='pie',ax=axes[1], title='Male', autopct='%1.1f%%', shadow=True)
plt.tight_layout()
plt.show()
Female
Male
group C
group D
group A
group E
group B
34.7%
24.9%
6.9%
13.3%
20.1%
28.8%
27.6%
11.0%
17.8%
14.7%
race/ethnicity
11:41 AM
1/16/2022

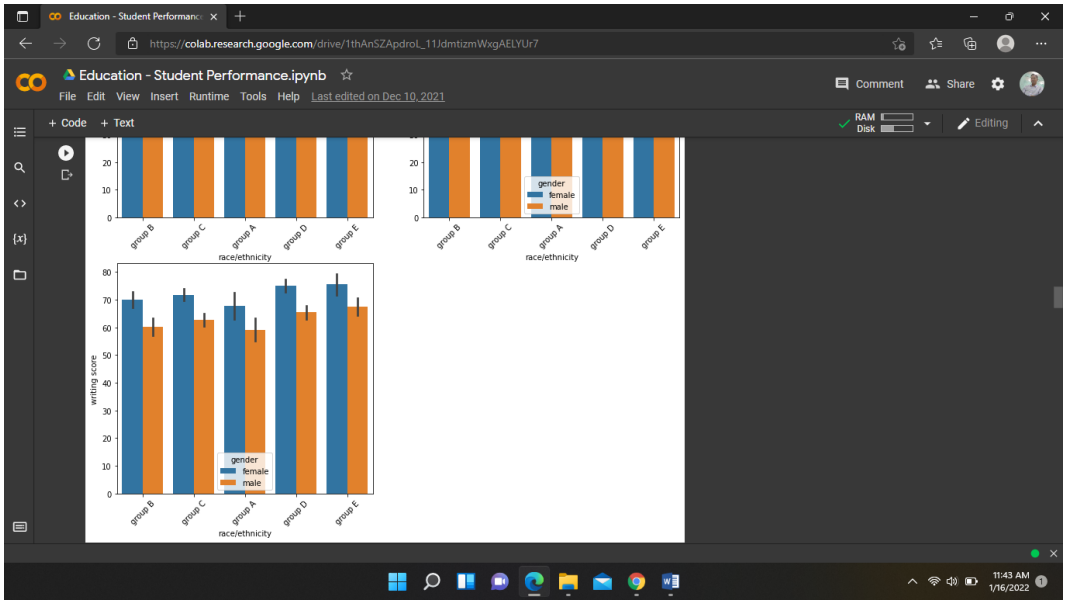

Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021
Comment Share
+ Code + Text
RAM Disk
Editing
Berdasarkan jenis kelamin pada kolom tingkat pendidikan orang tua
f, axes = plt.subplots(1,2, figsize=(15, 20))
female_data['parental level of education'].value_counts().plot(kind='pie',ax=axes[0], title='Female', autopct='%1.1f%%', shadow=True)
male_data['parental level of education'].value_counts().plot(kind='pie',ax=axes[1], title='Male', autopct='%1.1f%%', shadow=True)
plt.tight_layout()
plt.show()
Female
Male
associate's degree
some college
high school
master's degree
bachelor's degree
22.4%
22.8%
18.1%
6.9%
12.2%
17.6%
22.0%
22.4%
21.2%
4.8%
11.4%
18.3%
parental level of education
11:42 AM
1/16/2022

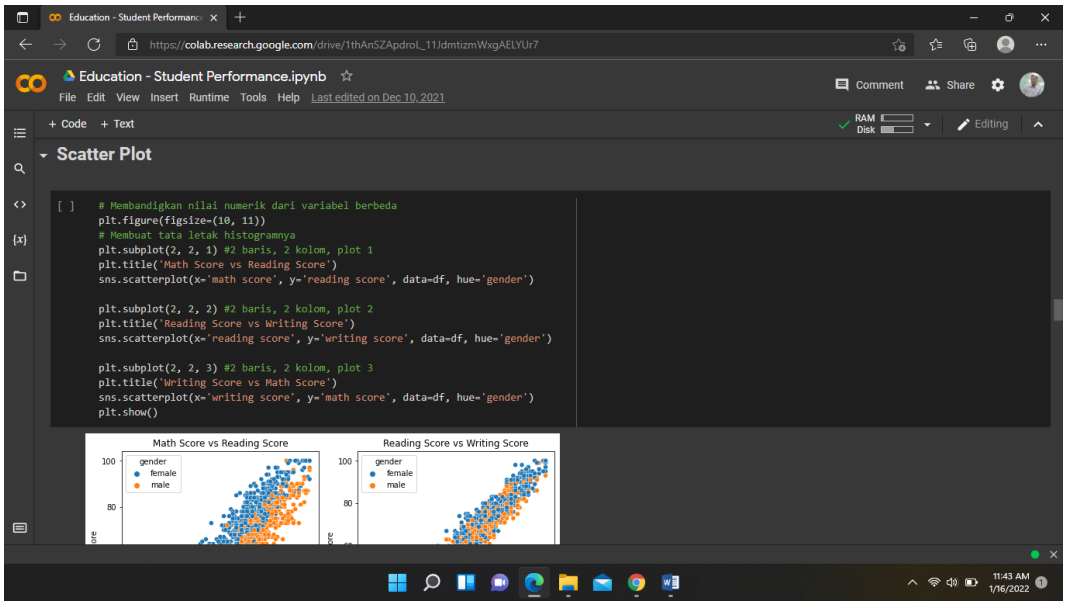

Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021
Comment
Share
+ Code + Text
RAM
Disk
Editing
Berdasarkan jenis kelamin pada kolom makan siang
f, axes = plt.subplots(1,2, figsize=(15, 20))
female_data['lunch'].value_counts().plot(kind='pie',ax=axes[0], title='Female', autopct='%1.1f%%', shadow=True)
male_data['lunch'].value_counts().plot(kind='pie',ax=axes[1], title='Male', autopct='%1.1f%%', shadow=True)
plt.tight_layout()
plt.show()
Female
Male
standard
63.5%
65.6%
lunch
11:42 AM
1/16/2022

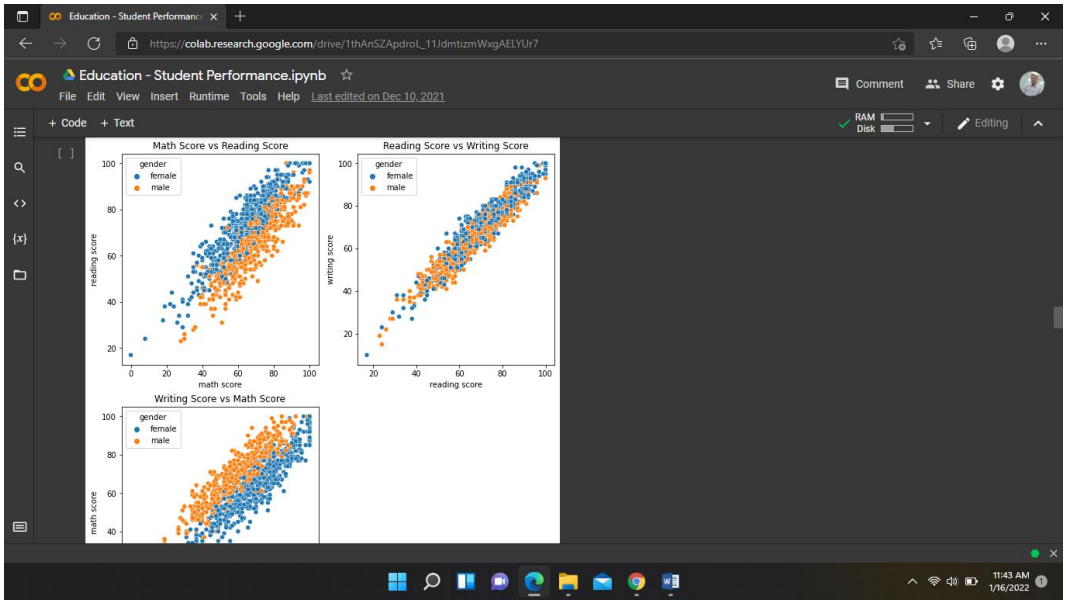

Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021
Comment
Share
+ Code + Text
RAM
Disk
Editing
Berdasarkan jenis kelamin pada kolom kursus persiapan ujian
f, axes = plt.subplots(1,2, figsize=(15, 20))
female_data['test preparation course'].value_counts().plot(kind='pie',ax=axes[0], title='Female', autopct='%1.1f%%', shadow=True)
male_data['test preparation course'].value_counts().plot(kind='pie',ax=axes[1], title='Male', autopct='%1.1f%%', shadow=True)
plt.tight_layout()
plt.show()
Female
Male
none
64.5%
63.9%
test preparation course
11:42 AM
1/16/2022

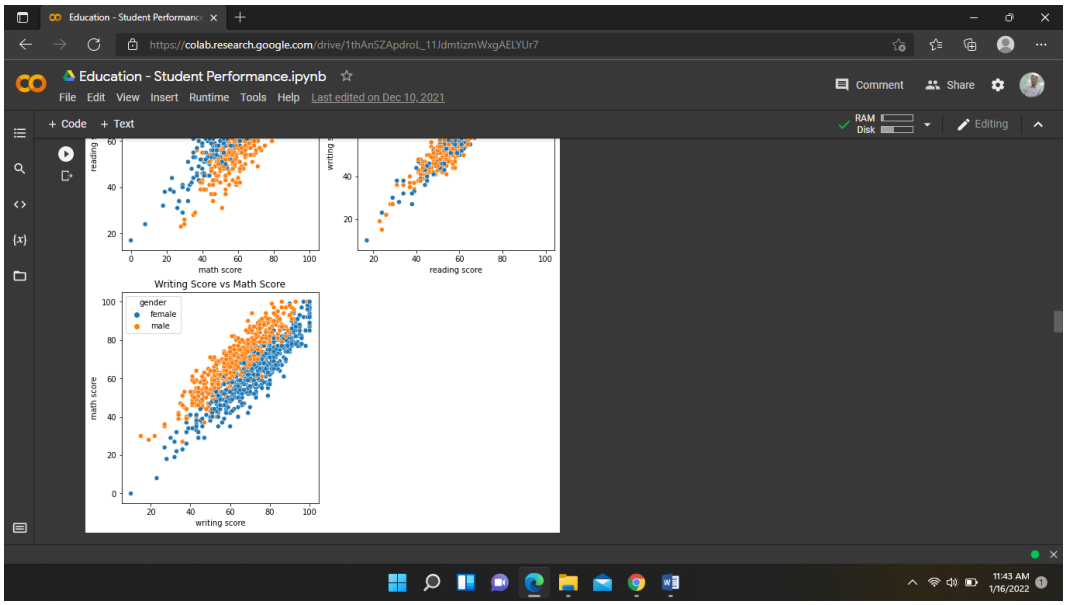

Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021
Comment
Share
+ Code + Text
RAM
Disk
Editing
Bar Chart
# Korelasi antar
# Berdasarkan kolom ras/etnis
plt.figure(figsize=(13, 12))
# Membuat tata letak histogramnya
plt.subplot(2, 2, 1) #2 baris, 2 kolom, plot 1
sns.barplot(x='race/ethnicity', y='math score',data=df, hue='gender') # Skor matematika
plt.xticks(rotation='45')
# Membuat tata letak histogramnya
plt.subplot(2, 2, 2) #2 baris, 2 kolom, plot 2
sns.barplot(x='race/ethnicity', y='reading score',data=df, hue='gender') # Skor membaca
plt.xticks(rotation='45')
# Membuat tata letak histogramnya
plt.subplot(2, 2, 3) #2 baris, 2 kolom, plot 3
sns.barplot(x='race/ethnicity', y='writing score',data=df, hue='gender') # Skor menulis
plt.xticks(rotation='45')
plt.show()
gender
female
male
11:43 AM
1/16/2022


Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021
Comment
Share
+ Code + Text
RAM
Disk
Editing
gender
female
male
math score
reading score
group B
group C
group A
group D
group E
race/ethnicity
11:43 AM
1/16/2022

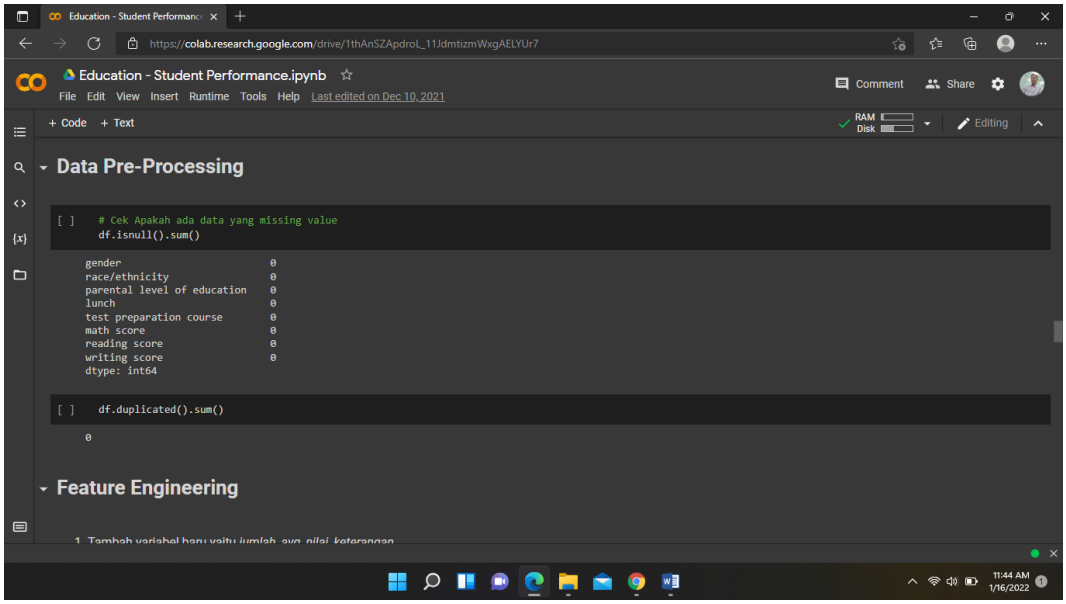

Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021
Comment
Share
+ Code
+ Text
RAM
Disk
Editing
group B
group C
group A
group D
group E
race/ethnicity
writing score
gender
female
male
11:43 AM
1/16/2022


Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021
Comment
Share
+ Code
+ Text
RAM
Disk
Editing
Scatter Plot
[ ]
# Membandigkan nilai numerik dari variabel berbeda
plt.figure(figsize=(10, 11))
# Membuat tata letak histogramnya
plt.subplot(2, 2, 1) #2 baris, 2 kolom, plot 1
plt.title('Math Score vs Reading Score')
sns.scatterplot(x='math score', y='reading score', data=df, hue='gender')
plt.subplot(2, 2, 2) #2 baris, 2 kolom, plot 2
plt.title('Reading Score vs Writing Score')
sns.scatterplot(x='reading score', y='writing score', data=df, hue='gender')
plt.subplot(2, 2, 3) #2 baris, 2 kolom, plot 3
plt.title('Writing Score vs Math Score')
sns.scatterplot(x='writing score', y='math score', data=df, hue='gender')
plt.show()
Math Score vs Reading Score
Reading Score vs Writing Score
gender
female
male
11:43 AM
1/16/2022

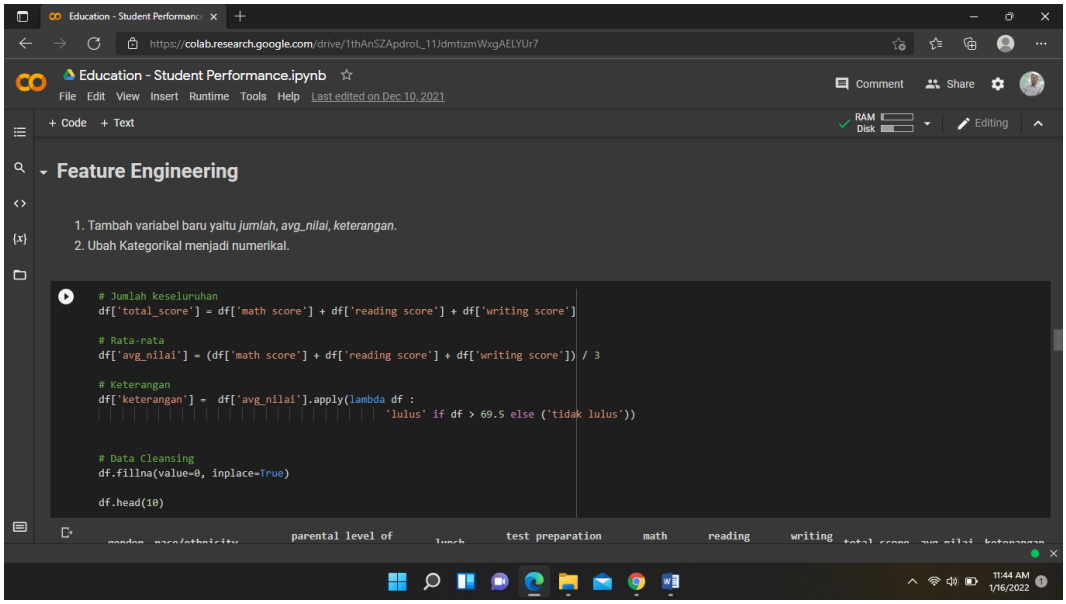

Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021
Comment
Share
+ Code
+ Text
RAM
Disk
Editing
Math Score vs Reading Score
Reading Score vs Writing Score
Writing Score vs Math Score
gender
female
male
reading score
writing score
math score
11:43 AM
1/16/2022

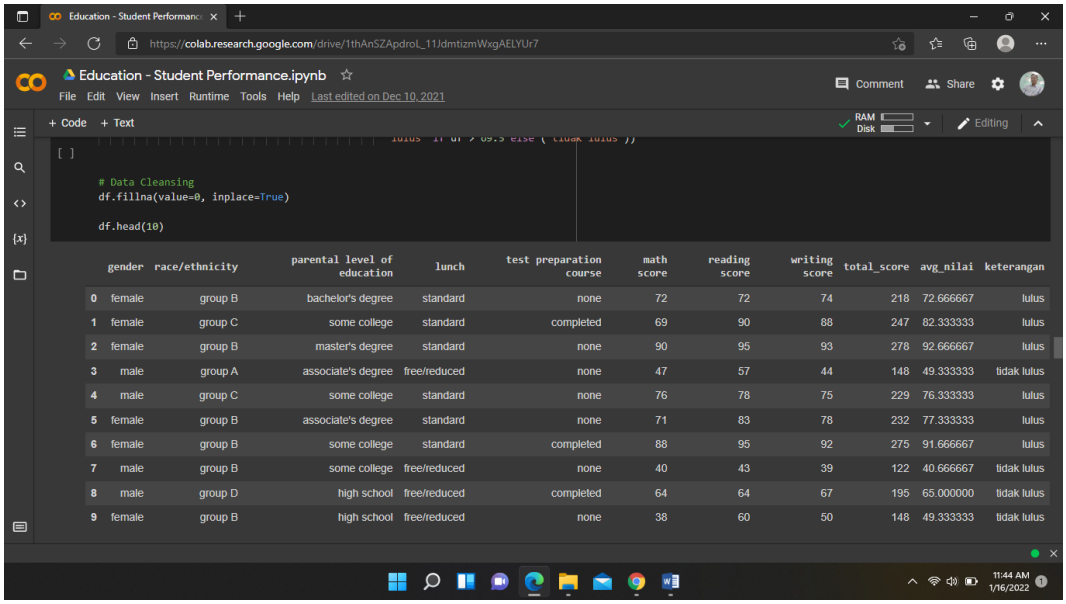

Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021
Comment
Share
+ Code
+ Text
RAM
Disk
Editing
Writing Score vs Math Score
gender
female
male
math score
writing score
11:43 AM
1/16/2022

Data Preprocessing

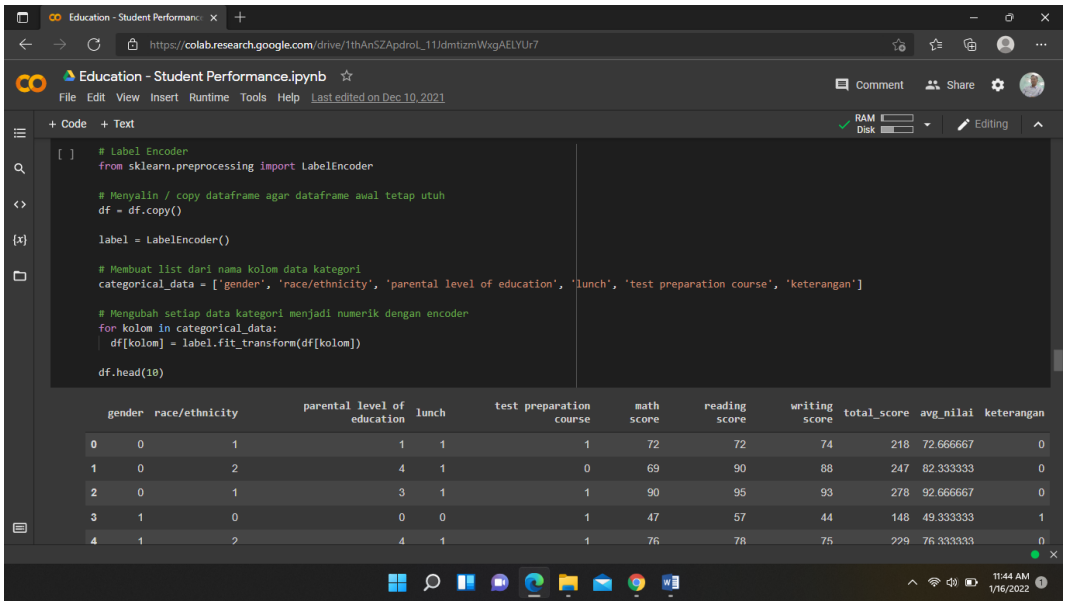


Feature Engineering

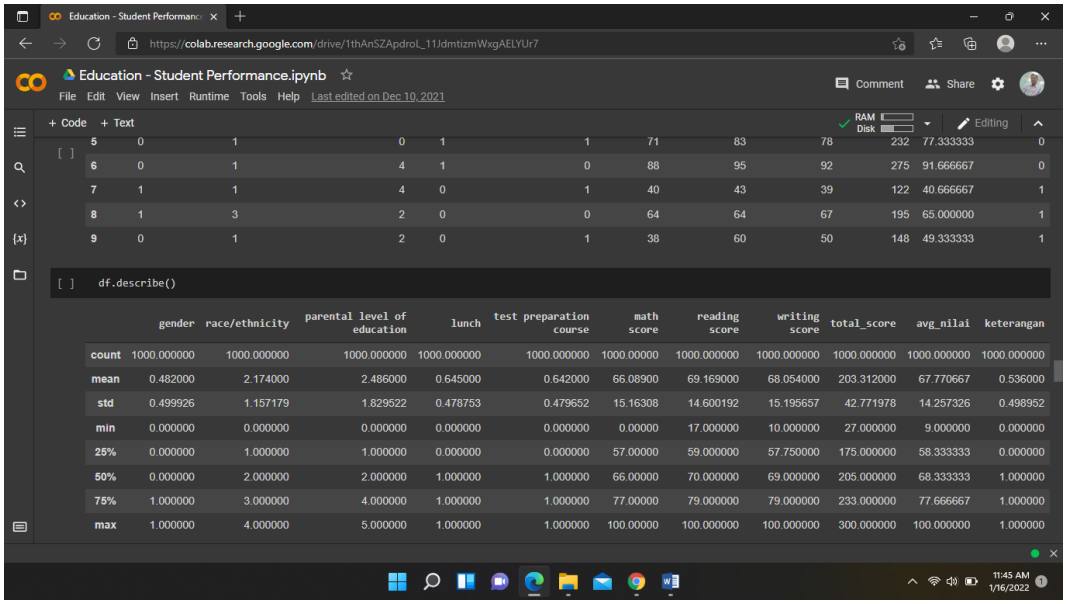




	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	total_score	avg_nilai	keterangan
0	female	group B	bachelor's degree	standard	none	72	72	74	218	72.666667	lulus
1	female	group C	some college	standard	completed	69	90	88	247	82.333333	lulus
2	female	group B	master's degree	standard	none	90	95	93	278	92.666667	lulus
3	male	group A	associate's degree	free/reduced	none	47	57	44	148	49.333333	tidak lulus
4	male	group C	some college	standard	none	76	78	75	229	76.333333	lulus
5	female	group B	associate's degree	standard	none	71	83	78	232	77.333333	lulus
6	female	group B	some college	standard	completed	88	95	92	275	91.666667	lulus
7	male	group B	some college	free/reduced	none	40	43	39	122	40.666667	tidak lulus
8	male	group D	high school	free/reduced	completed	64	64	67	195	65.000000	tidak lulus
9	female	group B	high school	free/reduced	none	38	60	50	148	49.333333	tidak lulus



	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	total_score	avg_nilai	keterangan
0	0	1	1	1	1	72	72	74	218	72.666667	0
1	0	2	4	1	0	69	90	88	247	82.333333	0
2	0	1	3	1	1	90	95	93	278	92.666667	0
3	1	0	0	0	1	47	57	44	148	49.333333	1
4	1	2	4	1	1	76	78	75	229	76.333333	0



Education - Student Performance.ipynb

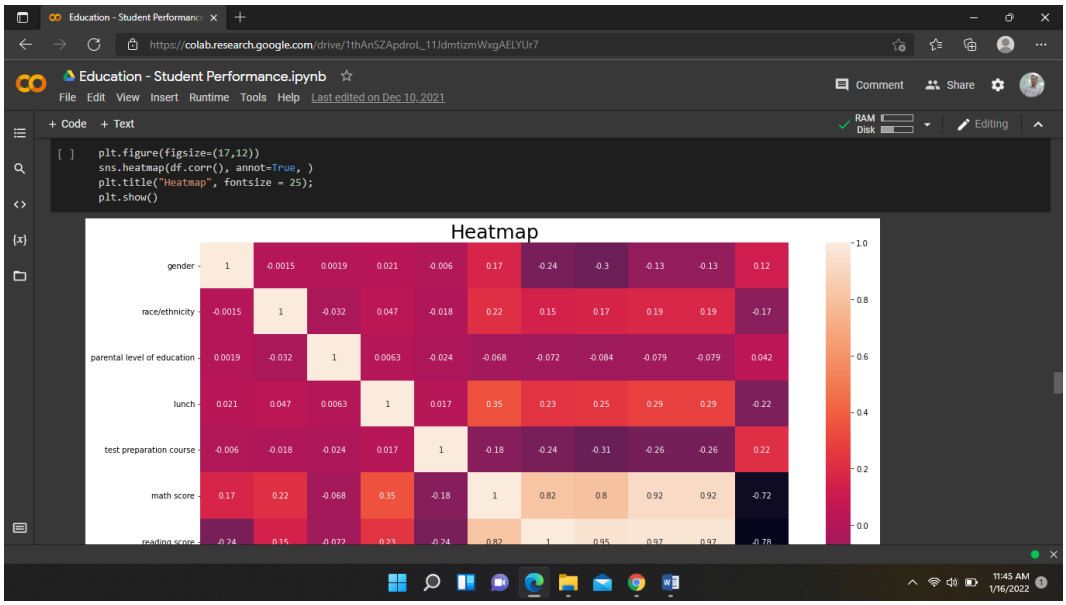
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021

+ Code + Text

5	0	1	0	1	1	71	83	78	232	77.333333	0
6	0	1	4	1	0	88	95	92	275	91.666667	0
7	1	1	4	0	1	40	43	39	122	40.666667	1
8	1	3	2	0	0	64	64	67	195	65.000000	1
9	0	1	2	0	1	38	60	50	148	49.333333	1

```
[ ]  df.describe()
```

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	total_score	avg_nilai	keterangan
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.00000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000
mean	0.482000	2.174000	2.486000	0.645000	0.642000	66.08900	69.169000	68.054000	203.312000	67.770667	0.536000
std	0.499926	1.157179	1.829522	0.478753	0.479652	15.16308	14.600192	15.195657	42.771978	14.257326	0.498952
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	17.000000	10.000000	27.000000	9.000000	0.000000
25%	0.000000	1.000000	1.000000	0.000000	0.000000	57.00000	59.000000	57.750000	175.000000	58.333333	0.000000
50%	0.000000	2.000000	2.000000	1.000000	1.000000	66.00000	70.000000	69.000000	205.000000	68.333333	1.000000
75%	1.000000	3.000000	4.000000	1.000000	1.000000	77.00000	79.000000	79.000000	233.000000	77.666667	1.000000
max	1.000000	4.000000	5.000000	1.000000	1.000000	100.00000	100.000000	100.000000	300.000000	100.000000	1.000000

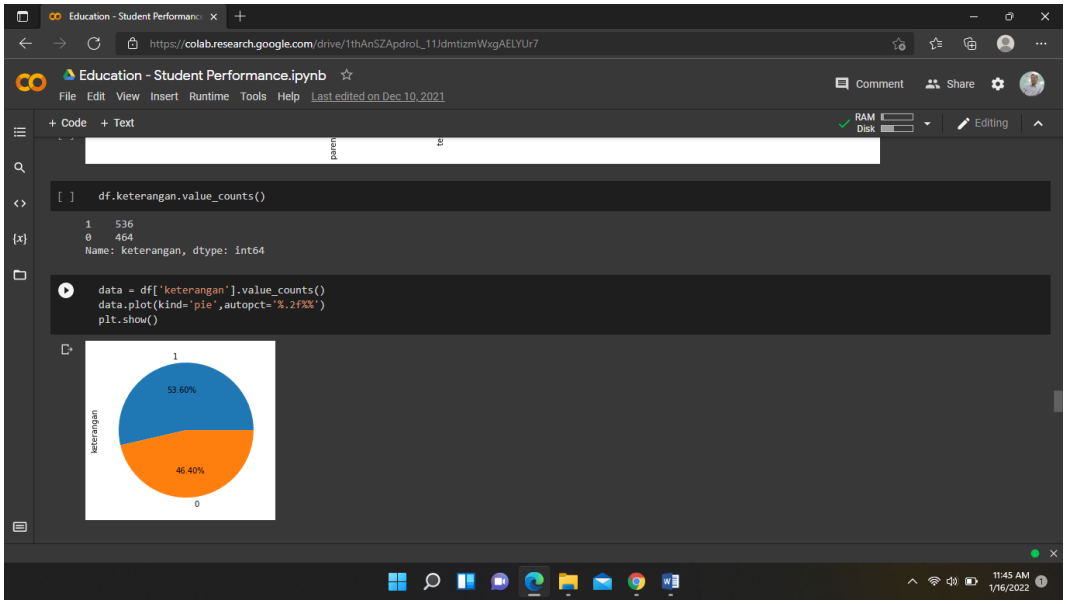


Education - Student Performance.ipynb

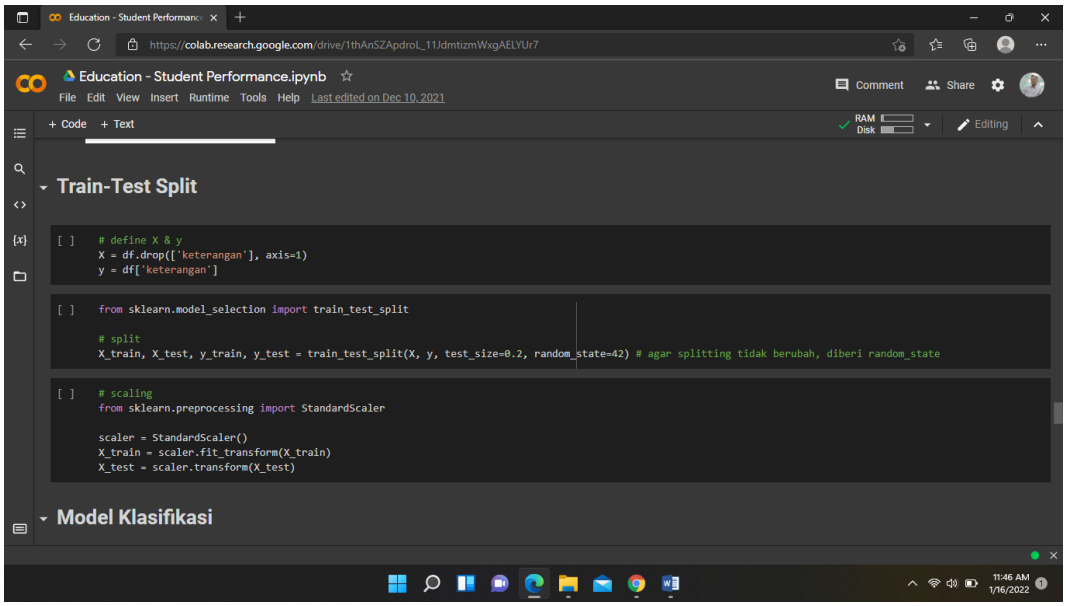
File Edit View Insert Runtime Tools Help Last edited on Dec 10, 2021

+ Code + Text

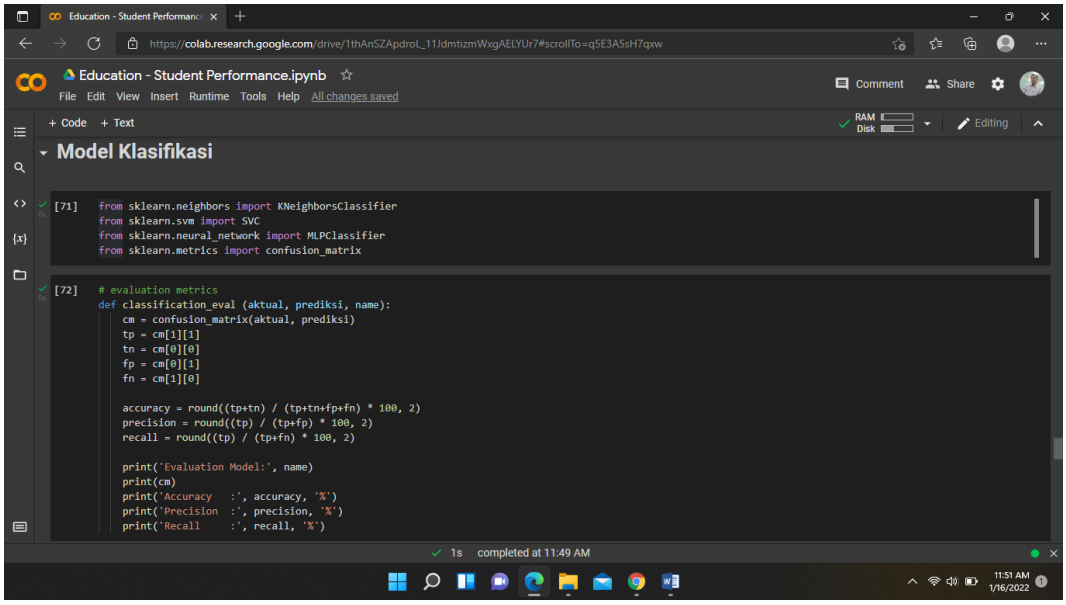
```
[ ]  plt.figure(figsize=(17,12))
     sns.heatmap(df.corr(), annot=True, )
     plt.title("Heatmap", fontsize = 25);
     plt.show()
```

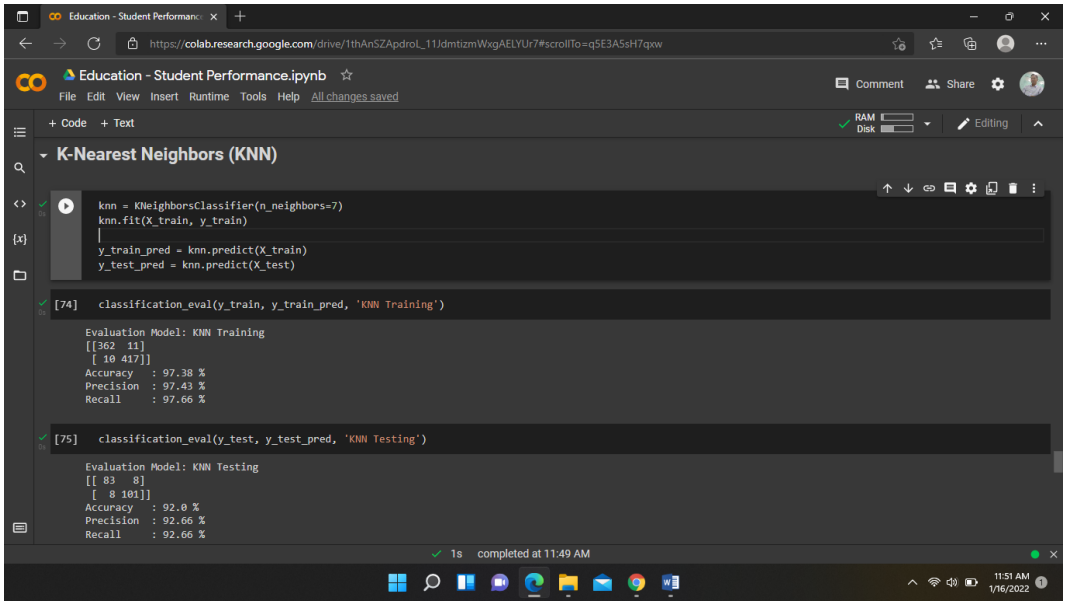


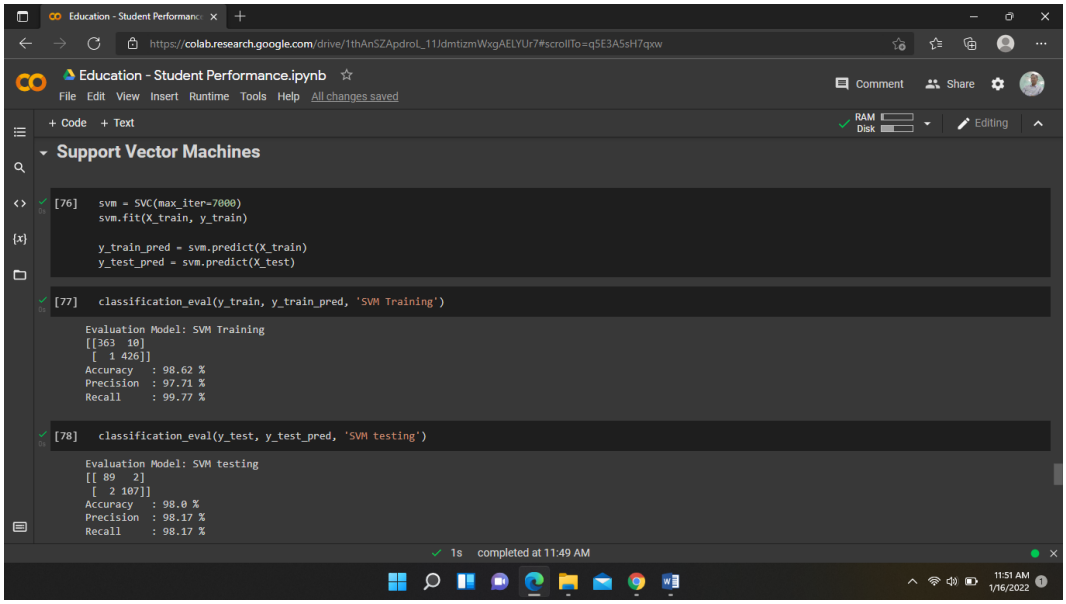
Train-Test Split

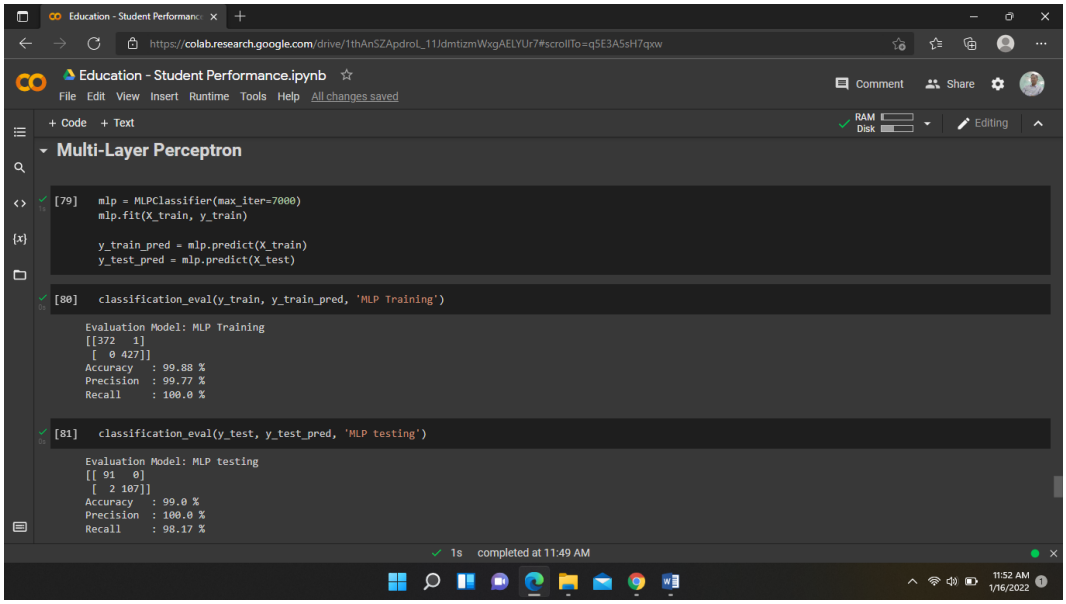


Evaluation Model Machine Learning

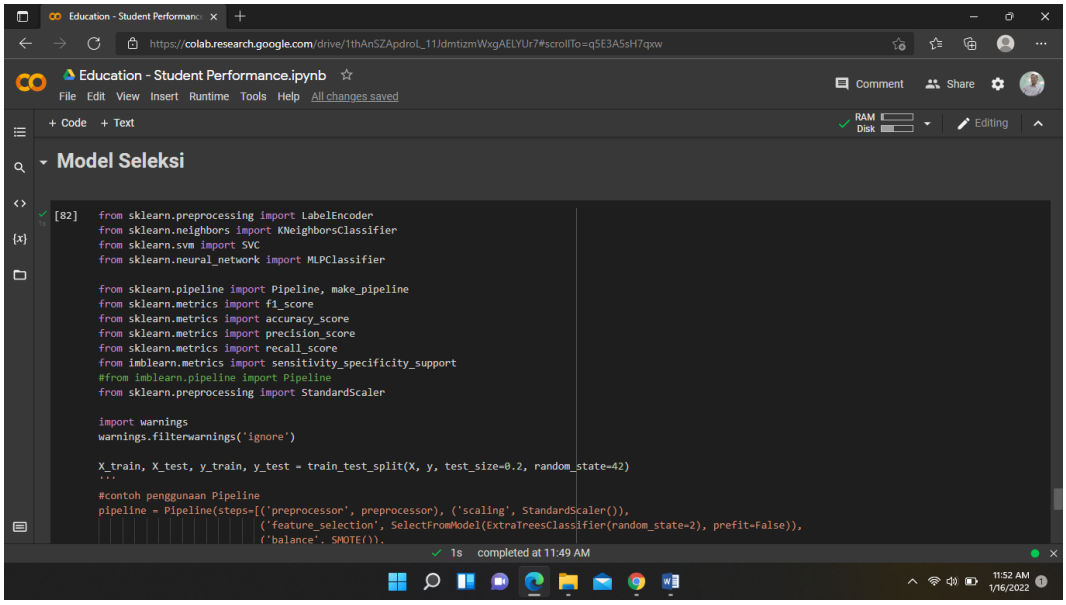


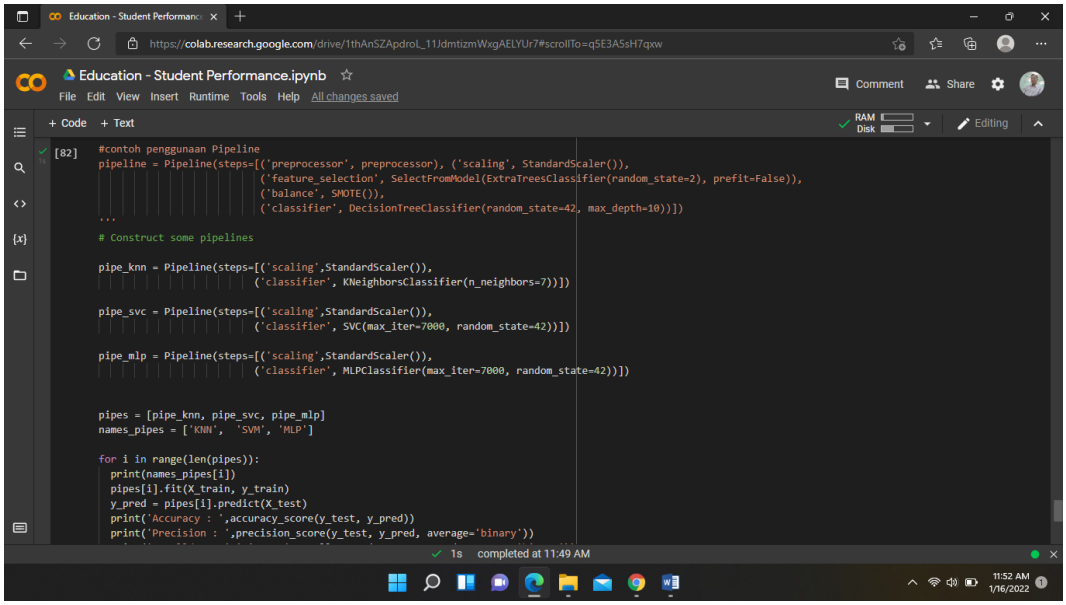


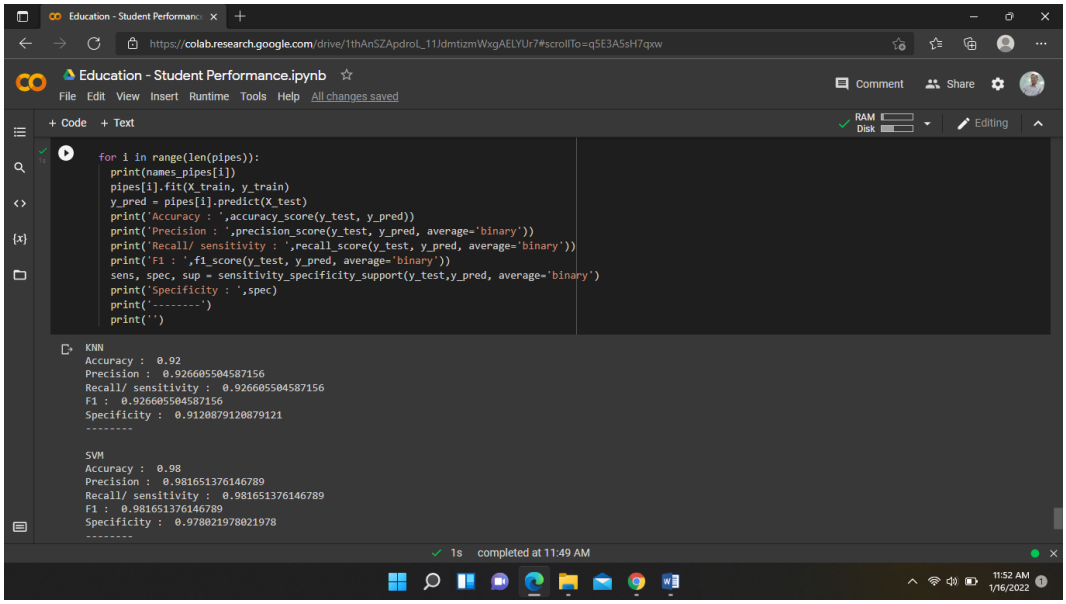

Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk
Editing
Support Vector Machines
[76] svm = SVC(max_iter=7000)
svm.fit(X_train, y_train)
y_train_pred = svm.predict(X_train)
y_test_pred = svm.predict(X_test)
[77] classification_eval(y_train, y_train_pred, 'SVM Training')
Evaluation Model: SVM Training
[[363 10]
[1 426]]
Accuracy : 98.62 %
Precision : 97.71 %
Recall : 99.77 %
[78] classification_eval(y_test, y_test_pred, 'SVM testing')
Evaluation Model: SVM testing
[[89 2]
[2 107]]
Accuracy : 98.0 %
Precision : 98.17 %
Recall : 98.17 %
1s completed at 11:49 AM
11:51 AM
1/16/2022


Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk
Editing
Multi-Layer Perceptron
[79] mlp = MLPClassifier(max_iter=7000)
mlp.fit(X_train, y_train)
y_train_pred = mlp.predict(X_train)
y_test_pred = mlp.predict(X_test)
[80] classification_eval(y_train, y_train_pred, 'MLP Training')
Evaluation Model: MLP Training
[[372 1]
[0 427]]
Accuracy : 99.88 %
Precision : 99.77 %
Recall : 100.0 %
[81] classification_eval(y_test, y_test_pred, 'MLP testing')
Evaluation Model: MLP testing
[[91 0]
[2 107]]
Accuracy : 99.0 %
Precision : 100.0 %
Recall : 98.17 %
1s completed at 11:49 AM
11:52 AM
1/16/2022

Model Selection Machine Learning Pipeline

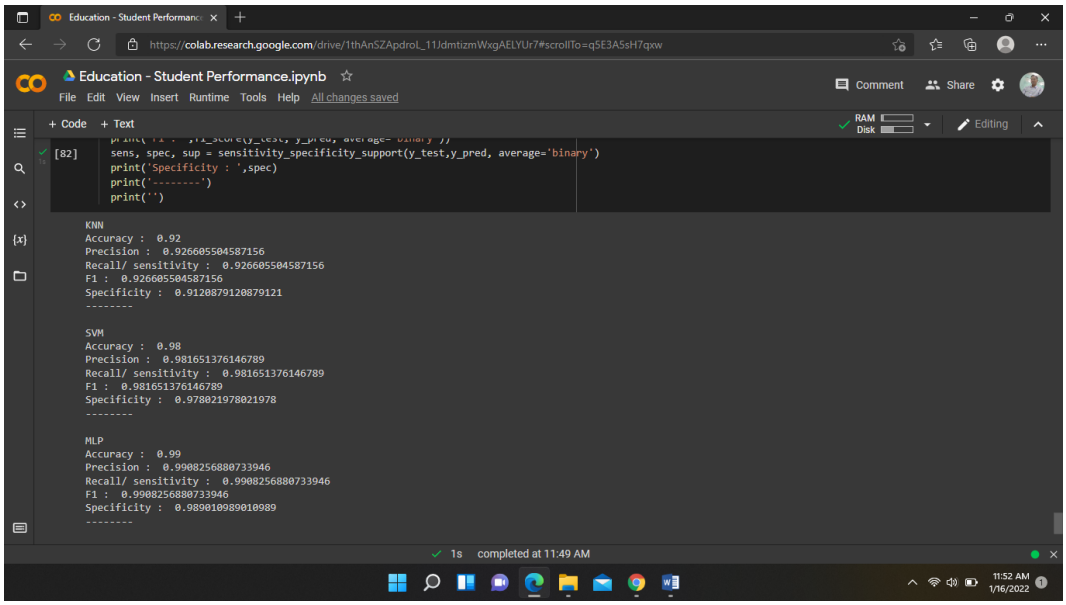





Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk
Editing
for i in range(len(pipes)):
print(names_pipes[i])
pipes[i].fit(X_train, y_train)
y_pred = pipes[i].predict(X_test)
print('Accuracy : ',accuracy_score(y_test, y_pred))
print('Precision : ',precision_score(y_test, y_pred, average='binary'))
print('Recall/ sensitivity : ',recall_score(y_test, y_pred, average='binary'))
print('F1 : ',f1_score(y_test, y_pred, average='binary'))
sens, spec, sup = sensitivity_specificity_support(y_test,y_pred, average='binary')
print('Specificity : ',spec)
print('--------')
print('')
KNN
Accuracy : 0.92
Precision : 0.926605504587156
Recall/ sensitivity : 0.926605504587156
F1 : 0.926605504587156
Specificity : 0.9120879120879121

SVM
Accuracy : 0.98
Precision : 0.981651376146789
Recall/ sensitivity : 0.981651376146789
F1 : 0.981651376146789
Specificity : 0.978021978021978

1s completed at 11:49 AM
11:52 AM
1/16/2022


Education - Student Performance.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment Share
+ Code + Text
RAM Disk
Editing
[82]
sens, spec, sup = sensitivity_specificity_support(y_test,y_pred, average='binary')
print('Specificity : ',spec)
print('--------')
print('')
KNN
Accuracy : 0.92
Precision : 0.926605504587156
Recall/ sensitivity : 0.926605504587156
F1 : 0.926605504587156
Specificity : 0.9120879120879121

SVM
Accuracy : 0.98
Precision : 0.981651376146789
Recall/ sensitivity : 0.981651376146789
F1 : 0.981651376146789
Specificity : 0.978021978021978

MLP
Accuracy : 0.99
Precision : 0.9908256880733946
Recall/ sensitivity : 0.9908256880733946
F1 : 0.9908256880733946
Specificity : 0.989010989010989

1s completed at 11:49 AM
11:52 AM
1/16/2022