



# PROJECT CASE PRESENTATION

JUDUL

PENERAPAN MACHINE LEARNING DALAM MEMPERSOALISASI PROFIL  
PENGGUNA BERDASARKAN AKTIVITAS PEMBELAJARAN ONLINE



Present by: ROYALS OUTLAWS ❤️



Rizal



Thesion



Rendika



# DAFTAR ISI

- 1 LATAR BELAKANG
- 2 TUJUAN, MANFAAT & KELUARAN
- 3 METODOLOGI PENGEMBANGAN
- 4 ANALISIS KEBUTUHAN
- 5 HASIL & PEMBAHASAN
- 6 KESIMPULAN
- 7 IMPLIKASI DAN REKOMENDASI
- 8 KETERBATASAN PENGERJAAN
- 9 REFERENCE



# LATAR BELAKANG



## Perkembangan Pembelajaran Online:

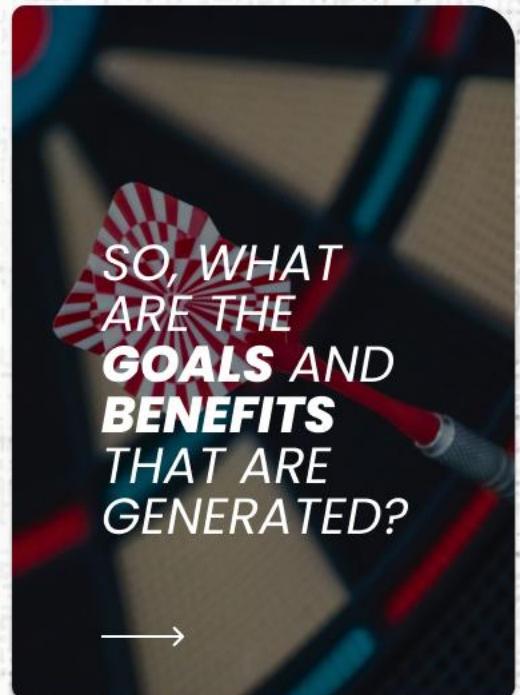
Platform pembelajaran online semakin banyak digunakan, tetapi tantangan utamanya adalah bagaimana menyediakan konten yang sesuai dengan kebutuhan dan kemampuan setiap pengguna.

## Pentingnya Personalisasi:

Personalisasi pembelajaran dapat meningkatkan efektivitas dan efisiensi belajar dengan memberikan konten yang relevan berdasarkan aktivitas dan preferensi pengguna. Penelitian menunjukkan bahwa personalisasi meningkatkan keterlibatan serta hasil belajar siswa.

## Pemanfaatan Machine Learning:

Teknologi machine learning memungkinkan prediksi profil pengguna berdasarkan aktivitas belajar mereka, sehingga konten pembelajaran dapat disesuaikan secara otomatis, meningkatkan pengalaman belajar secara keseluruhan.



# TUJUAN & MANFAAT

DIKEMBANGKANNYA MODEL DAN ANALISIS.



## MANFAAT :

- 1. Bagi Pengguna:** Pengalaman belajar yang lebih dipersonalisasi, relevan, dan efektif, yang dapat meningkatkan hasil akademik.
- 2. Bagi Pengelola Platform:** Dapat mengembangkan fitur personalisasi yang lebih canggih dan meningkatkan kualitas kurikulum.
- 3. Bagi Dunia Pendidikan:** Mendorong adopsi teknologi machine learning dalam pembelajaran online untuk meningkatkan inovasi pendidikan.



## TUJUAN :

1. Mengidentifikasi **faktor-faktor penting** yang mempengaruhi aktivitas dan kinerja belajar pengguna.
2. **Membangun model** machine learning untuk **memprediksi profil pengguna** berdasarkan data aktivitas mereka.
3. **Menyediakan wawasan** bagi pengelola platform untuk **mengoptimalkan kurikulum** sesuai kebutuhan pengguna.



## KELUARAN YANG DIHARAPKAN :

- 1. Model Machine Learning:** Model prediktif untuk personalisasi pembelajaran yang akurat dan sesuai dengan profil pengguna.
- 2. Peningkatan Efisiensi dan Kualitas Belajar:** Pengguna dapat mencapai tujuan pembelajaran lebih cepat dengan konten yang lebih relevan.
- 3. Rekomendasi Strategis:** Wawasan yang membantu pengelola platform dalam menyusun kurikulum dan fitur pembelajaran yang lebih sesuai dengan kebutuhan pengguna.
- 4. Peningkatan Motivasi dan Keterlibatan:** Personalisasi yang lebih baik akan meningkatkan motivasi belajar dan keterlibatan pengguna dengan platform.



# METODOLOGI PENELITIAN

## 6 Steps in **CRISP-DM**

The Standard  
Data Mining  
Process

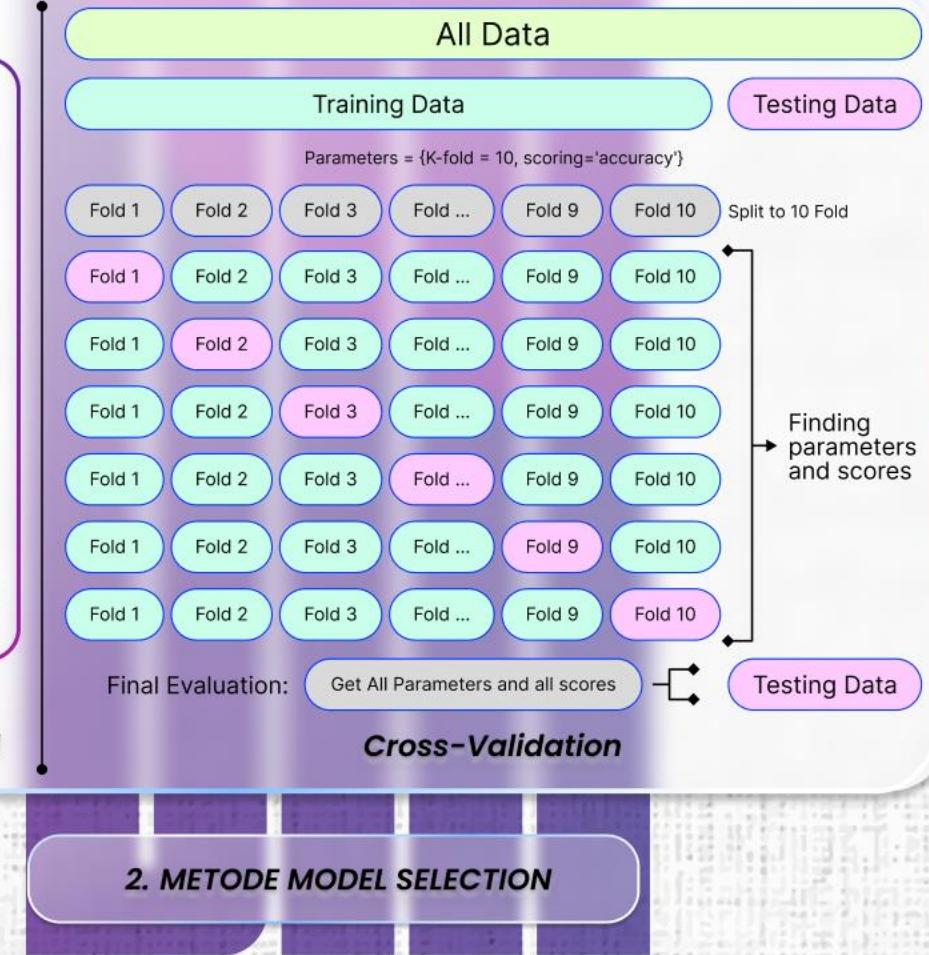
Credit : google.com

### 1. METODE ALUR PROSES Pengerjaan



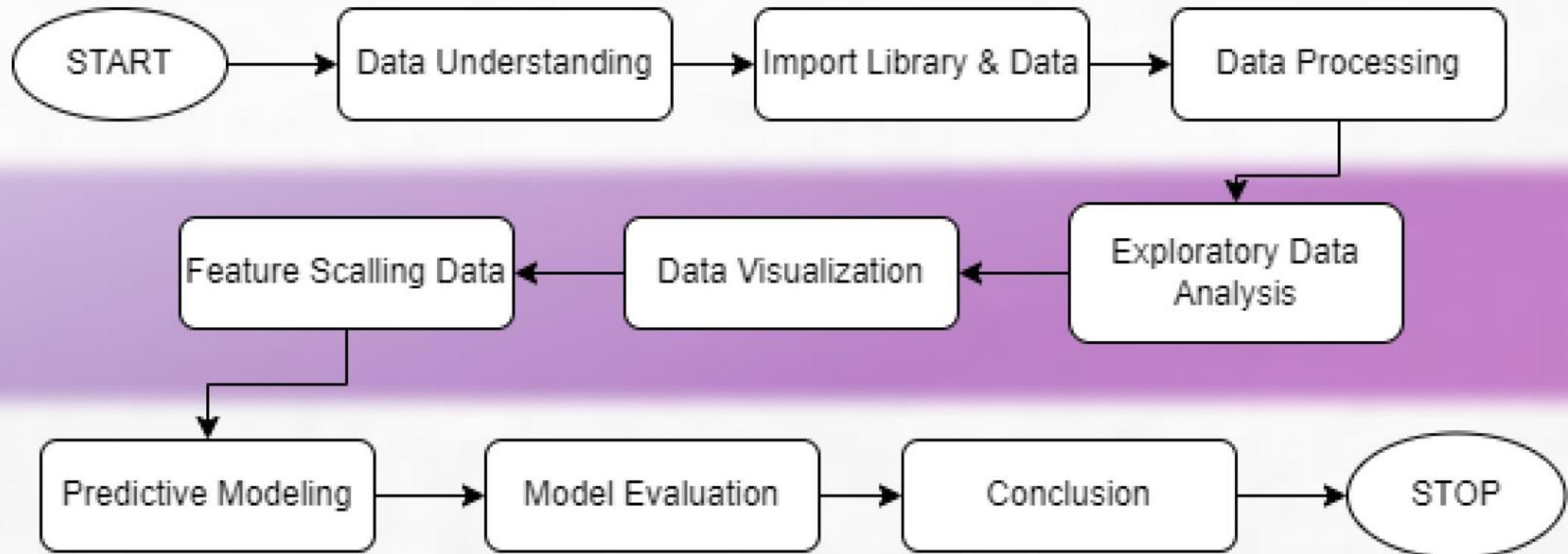
- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC

### Evaluasi Model



# METODOLOGI PENELITIAN

## 2. TAHAPAN PENGEMBANGAN



# ANALISIS KEBUTUHAN



## KEBUTUHAN INFRASTRUKTUR

### 1. Perangkat Keras:

- **Prosesor:** Intel Core i5-10300H CPU @ 2.50GHz (8CPUs)
- **RAM:** 16GB
- **GPU:** NVIDIA GeForce RTX 2060 with Max-Q Design

### 2. Perangkat Lunak:

- OS: Windows 11 x64
- Bahasa Pemrograman: Python
- Library Utama:
  - NumPy, Pandas, Matplotlib, Scikit-learn, Pingouin
- IDE/Editor: Jupyter Notebook via Google Colab

### 3. Lingkungan Pengembangan:

- Virtual Environment: Google Colab
- Penyimpanan: Google Drive untuk dataset dan backup hasil eksperimen.



## SUMBER DATA

### 1. DATASET

Disediakan oleh Innovative Competition and National Informatics Conference (ICONIC-IT)

### 2. PENGGUNAAN DATA

- Melatih model machine learning untuk memprediksi profil pengguna.
- Proses menggunakan Data Life Cycle dan prinsip Data Governance untuk memastikan kualitas dan konsistensi data.



## DATA YANG DIPERLUKAN

Data yang diperlukan meliputi identitas pengguna (NAME, USER\_ID), aktivitas belajar seperti jam belajar per bidang (HOURS\_DATASCIENCE, HOURS\_BACKEND, HOURS\_FRONTEND), jumlah kursus pemula dan lanjutan (NUM\_COURSES\_BEGINNER\_DATASCIENCE, NUM\_COURSES\_ADVANCED\_DATASCIENCE, dan seterusnya), serta skor rata-rata (AVG\_SCORE\_DATASCIENCE, AVG\_SCORE\_BACKEND, AVG\_SCORE\_FRONTEND). Target prediksi adalah Profil Pengguna (PROFILE).



# HASIL DAN PEMBAHASAN

## 1. Data Processing

### a) Handling Missing Value

1. Total **685 nilai null** ditemukan di beberapa kolom.
2. **Imputasi missing value** dilakukan dengan **metode mean** untuk kolom numerik, dipisahkan **berdasarkan kelas** label 'PROFILE' bertujuan untuk **menjaga representasi data** yang akurat

Tabel 1 Hasil pemeriksaan sebelum handling missing value

VARIABLE	COUNT OF NULL
NAME	0
USER_ID	0
HOURS_DATASCIENCE	14
HOURS_BACKEND	53
HOURS_FRONTEND	16
NUM_COURSES_BEGINNER_DATASCIENCE	26
NUM_COURSES_BEGINNER_BACKEND	18
NUM_COURSES_BEGINNER_FRONTEND	39
NUM_COURSES_ADVANCED_DATASCIENCE	2
NUM_COURSES_ADVANCED_BACKEND	8
NUM_COURSES_ADVANCED_FRONTEND	37
AVG_SCORE_DATASCIENCE	220
AVG_SCORE_BACKEND	84
AVG_SCORE_FRONTEND	168
PROFILE	0

### B) Handling Outlier

1. **Metode IQR** digunakan untuk **mendeteksi outlier**, dengan batasan kuartil **Q1 dan Q3**.
2. terdapat **total 1006 baris outlier**, yang meskipun relatif kecil dibandingkan dengan **jumlah total dataset sebesar 20000 baris**, tetap **perlu dibuang** karena dapat mengganggu analisis. Setelah **outlier dihapus**, dataset menjadi lebih bersih, dengan total data akhir **sebanyak 19035 baris**.

Tabel 2 Hasil pemeriksaan sebelum handling outlier

NUMERICAL VARIABLE	COUNT OF OUTLIER
HOURS_DATASCIENCE	71
HOURS_BACKEND	84
HOURS_FRONTEND	49
NUM_COURSES_BEGINNER_DATASCIENCE	71
NUM_COURSES_BEGINNER_BACKEND	117
NUM_COURSES_BEGINNER_FRONTEND	5
NUM_COURSES_ADVANCED_DATASCIENCE	57
NUM_COURSES_ADVANCED_BACKEND	72
NUM_COURSES_ADVANCED_FRONTEND	132
AVG_SCORE_DATASCIENCE	105
AVG_SCORE_BACKEND	114
AVG_SCORE_FRONTEND	129



# HASIL DAN PEMBAHASAN



## 1. Data Processing

### c) Feature Scaling

1. Digunakan untuk **skala ulang data(Feature Scaling)** numerik ke **rentang 0 hingga 1** dengan metode **Min-max Scaler**.
2. **Alasan** menggunakan Feature Scaling karena **rentang data pada dataset memang berbeda** antara variabel numerik satu dengan yang lainnya, dan ingin **membantu mempercepat proses pembelajaran** pada machine learning
3. **Min-max** dapat dinyatakan dalam persamaan berikut:

$$W_{\text{norm}} = \frac{W_i - W_{\min}}{W_{\max} - W_{\min}}$$

Tabel 3 Contoh hasil dari sebelum dan sesudah normalisasi min-max

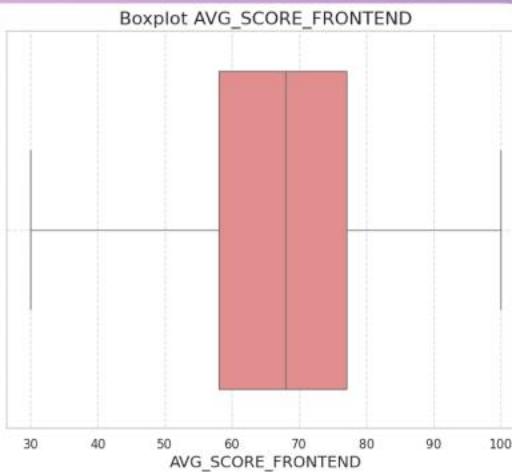
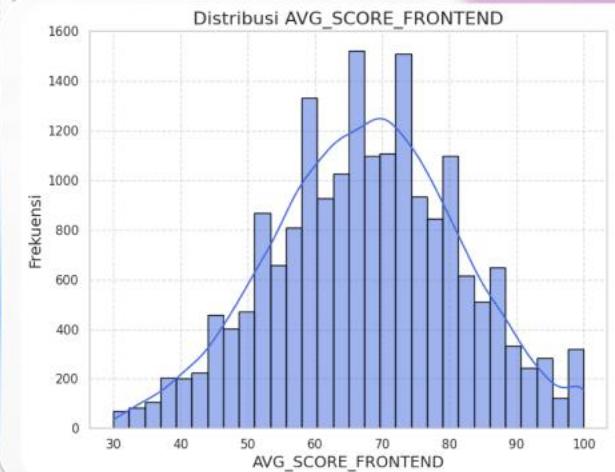
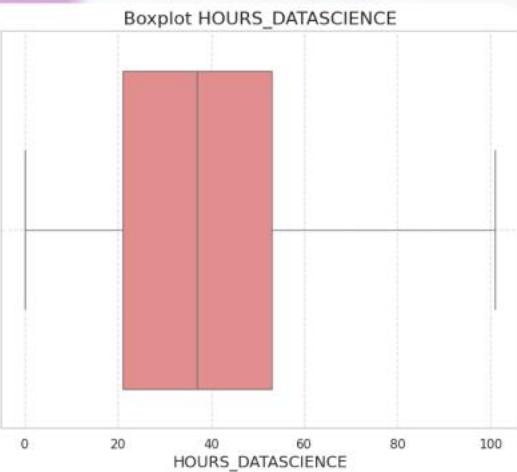
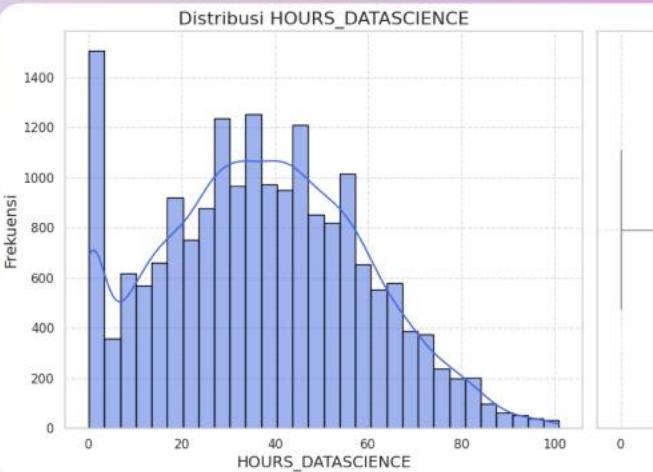
Sebelum		Sesudah	
HOURS_D ATASCIEN	NUM_COURSES_A DVANCED_BACK	HOURS_D ATASCIEN	NUM_COURSES_A DVANCED_BACK
CE	END	CE	END
37	1	0,366337	0,1
16	6	0,158416	0,6
...	...	...	...
38	5	0,376238	0,5
30	5	0,29703	0,5



# HASIL DAN PEMBAHASAN



## 2. Exploratory Data Analysis



### 1. DISTRIBUSI BEBERAPA VARIABEL



Present by: ROYALS OUTLAWS

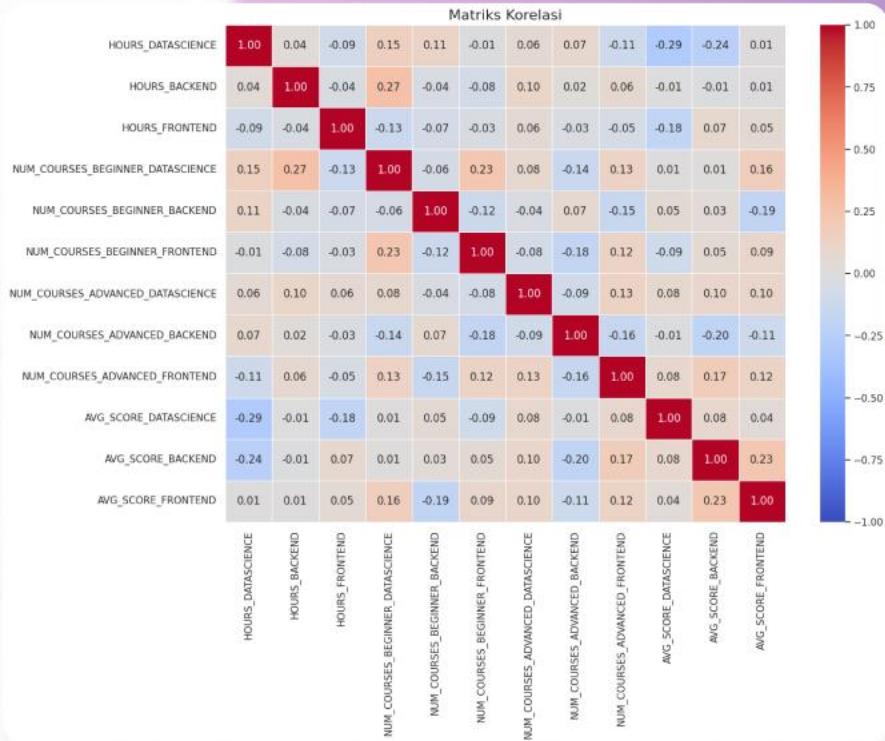
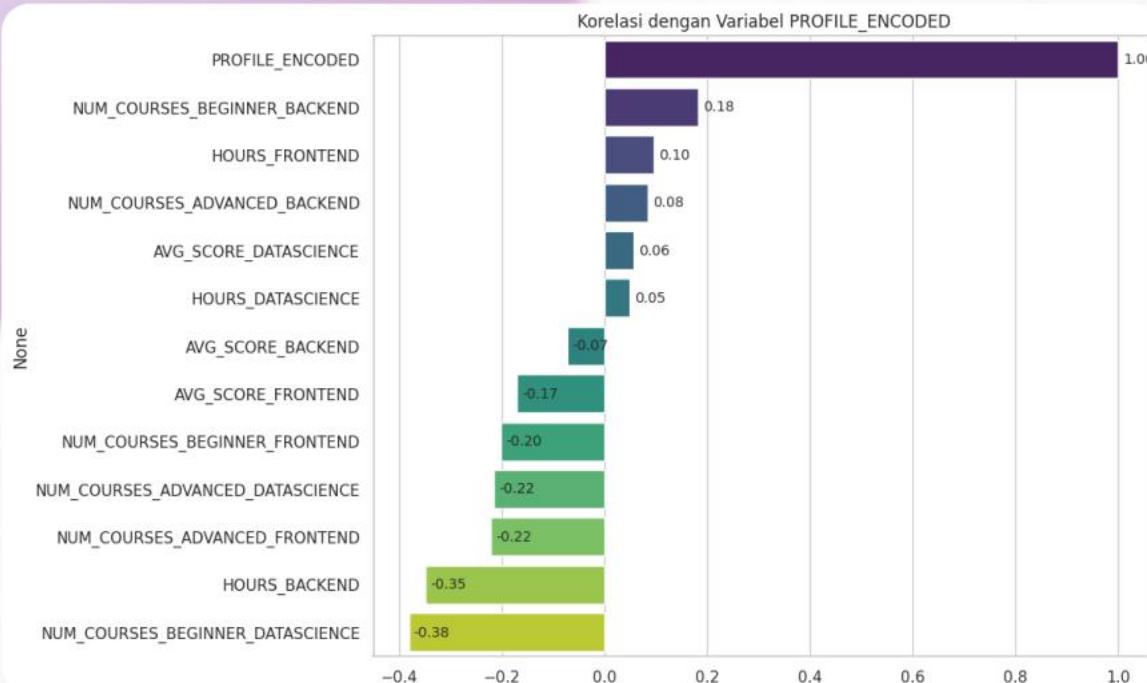
Slide No

10

# HASIL DAN PEMBAHASAN



## 2. Exploratory Data Analysis



## 2. KORELASI SEMUA VARIABEL DENGAN VARIABEL TARGET

## 3. KORELASI ANTAR VARIABEL



Present by: ROYALS OUTLAWS

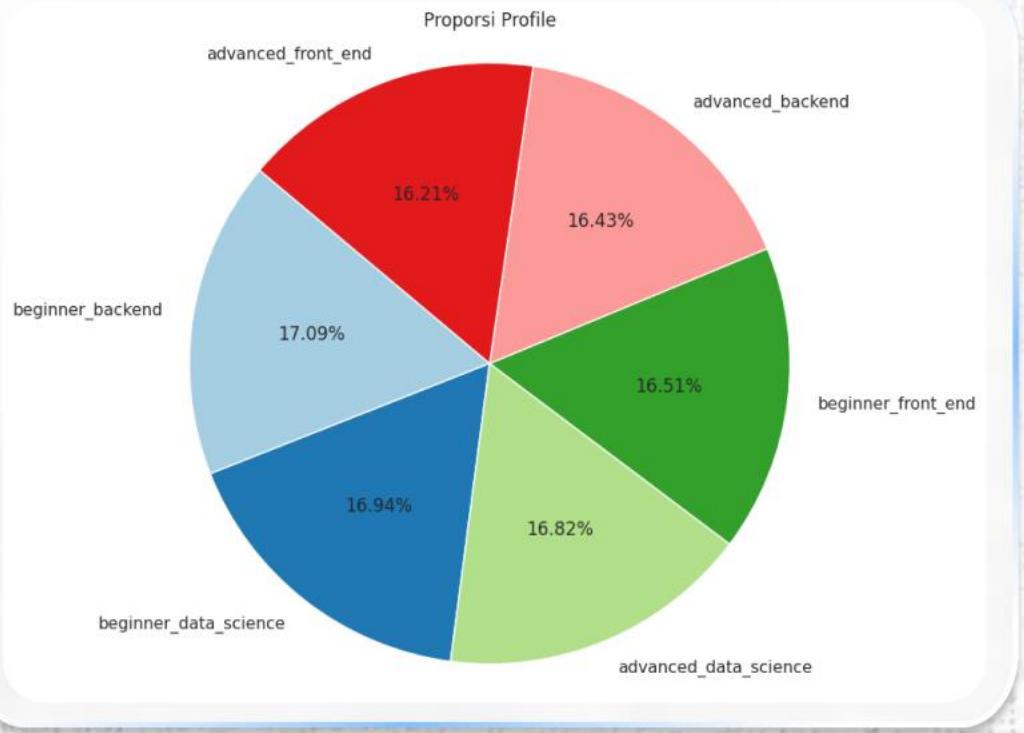
Slide No 11

# HASIL DAN PEMBAHASAN

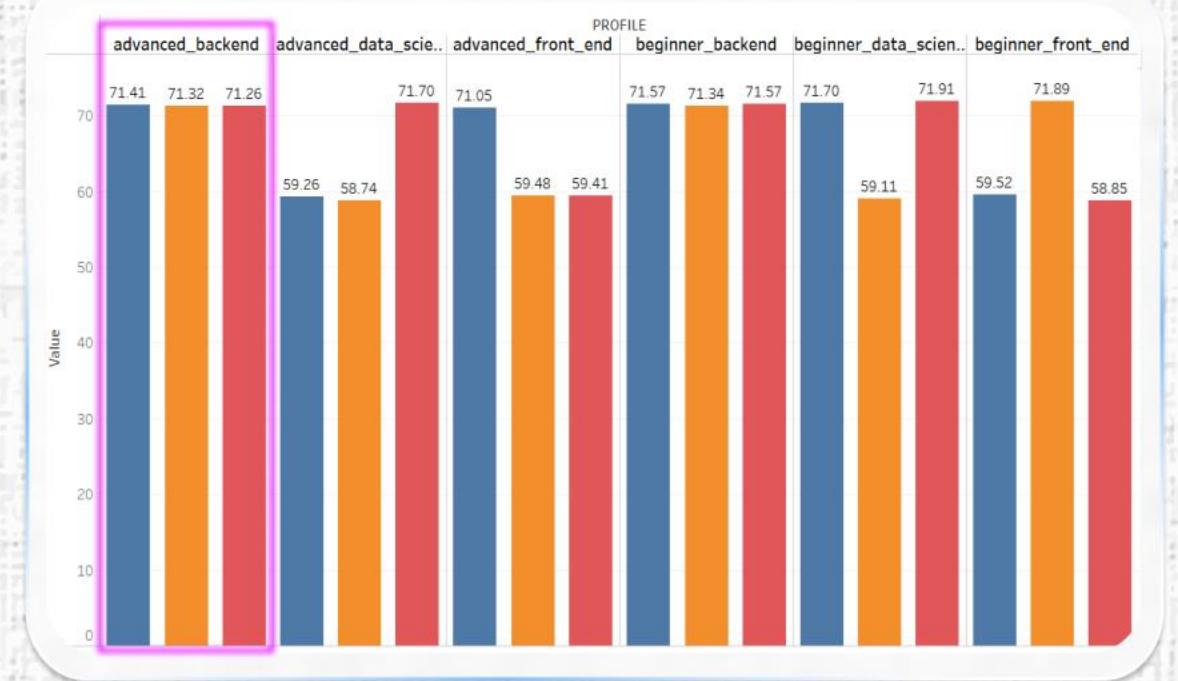


## 3. WAWASAN BERHARGA

- Avg. Hours\_Backend
- Avg. Hours\_DataScience
- Avg. Hours\_Frontend



1. GRAFIK PROPORSI PROFILE



2. PERBANDINGAN RERATA NILAI TIAP COURSE TIAP PROFILE



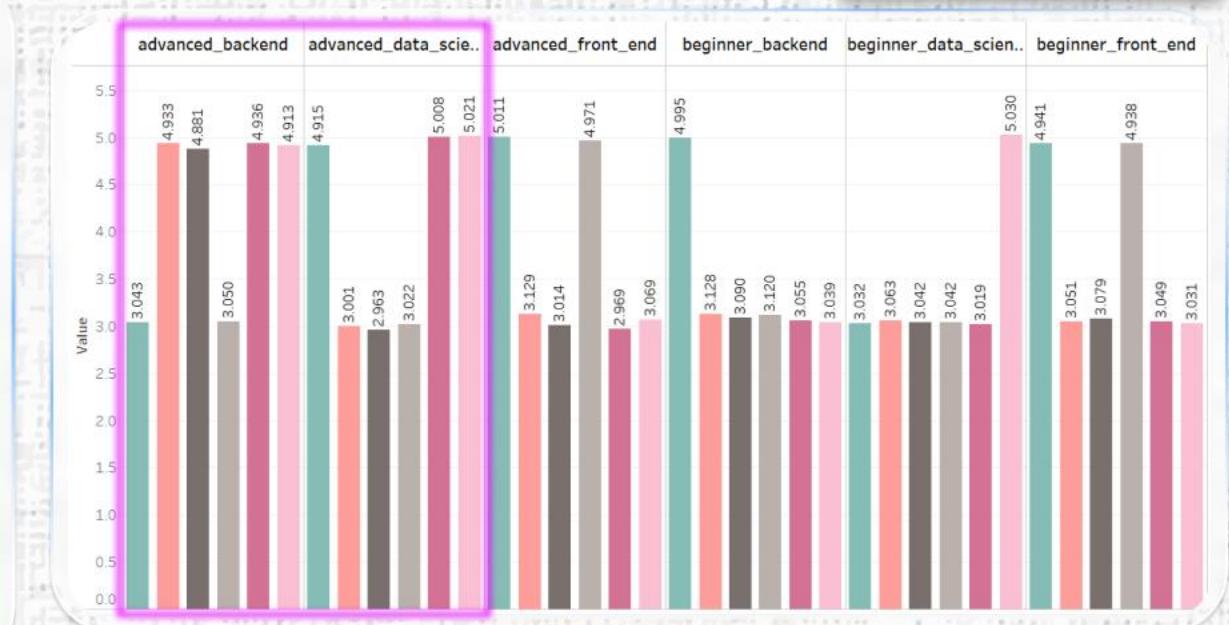
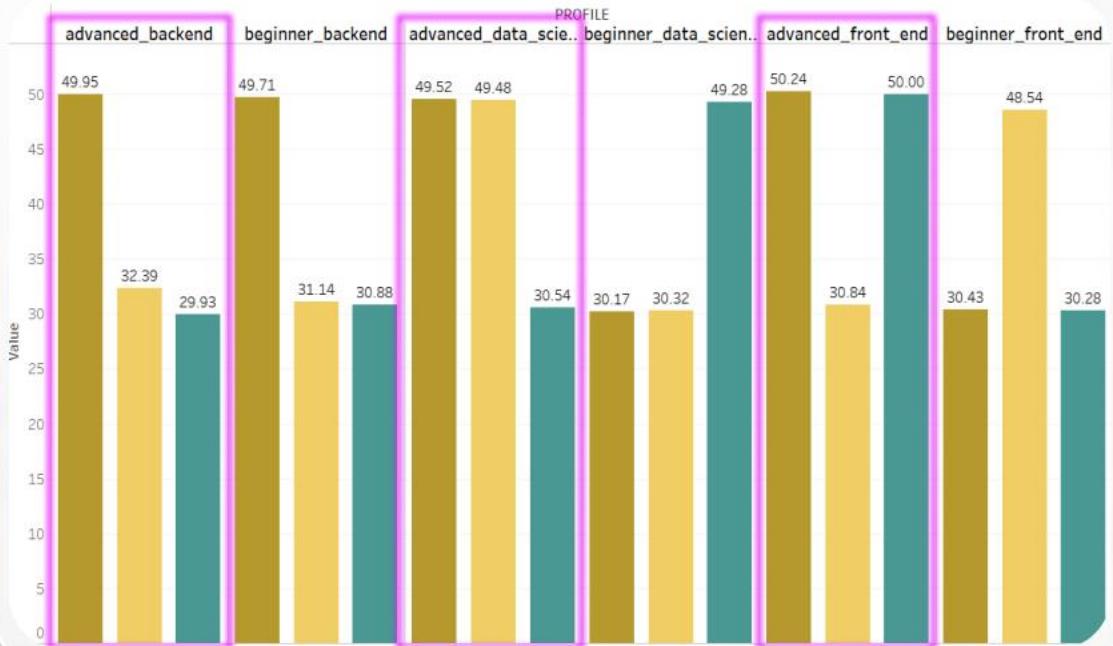
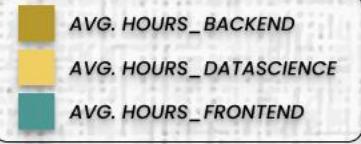
Present by: ROYALS OUTLAWS

Slide No 12

# HASIL DAN PEMBAHASAN



## 3. WAWASAN BERHARGA



## 4. PERBANDINGAN RERATA JUMLAH COURSE YANG DIIKUTI TIAP PROFILE

### 3. RERATA WAKTU YANG DIHABISKAN TIAP PROFILE SAAT MENERJAKAN COURSE



Present by: ROYALS OUTLAWS

Slide No 13

# HASIL DAN PEMBAHASAN

## 4. Pemilihan dan Pelatihan Model



### A) TAHAP PRE-MODELING

**Stratified Splitting:** Memastikan keseimbangan label antar kelas menggunakan stratifikasi



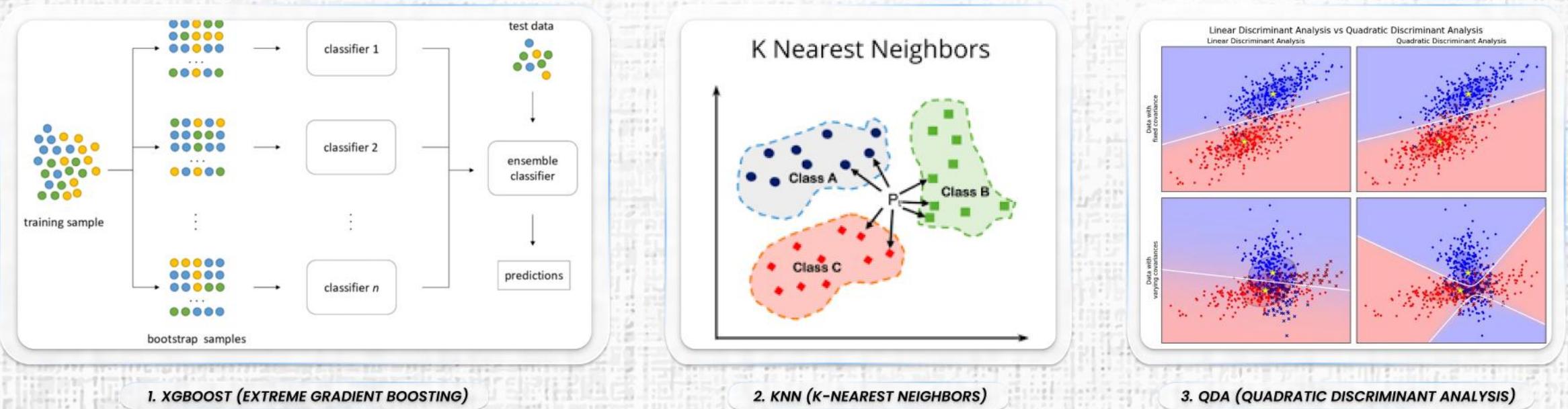
# HASIL DAN PEMBAHASAN



## 4. Pemilihan dan Pelatihan Model

### B) MODEL SELECTION

Pemilihan model melibatkan tiga model utama, yaitu QDA (quadratic discriminant analysis), XGBoost (eXtreme Gradient Boosting), dan KNN(k-nearest neighbors). Ketiga model ini dipilih setelah melalui berbagai trial and error serta evaluasi kinerja terhadap model-model lain, seperti Random Forest, Decision Tree, Logistic Regression, Support Vector Machine, LightGBM, dan lainnya. Keputusan untuk menggunakan ketiga model ini didasarkan pada kinerja superior mereka dalam menyelesaikan masalah yang dihadapi dibandingkan dengan model-model lainnya.

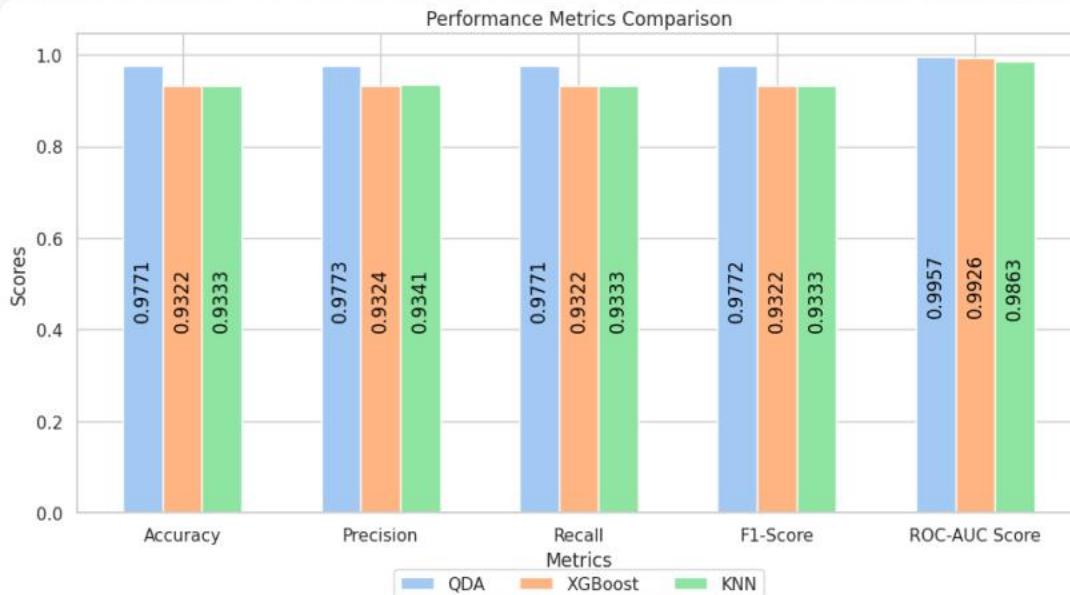


# HASIL DAN PEMBAHASAN

## 4. Pemilihan dan Pelatihan Model

### c) EVALUASI KINERJA

Evaluasi dilakukan dengan menguji ketahanan model terhadap variasi data input, menggunakan 20% data testing (3.807 baris). Karena data ini lebih cocok untuk klasifikasi, metrik evaluasi yang digunakan meliputi Accuracy, Precision, Recall, F1-score, dan ROC AUC.



$$\text{Accuracy} = \left( \frac{TP + TN}{TP + FP + FN + TN} \right)$$

$$F1 - score = \left( 2 \times \frac{Recall \times Precision}{Recall + Precision} \right)$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

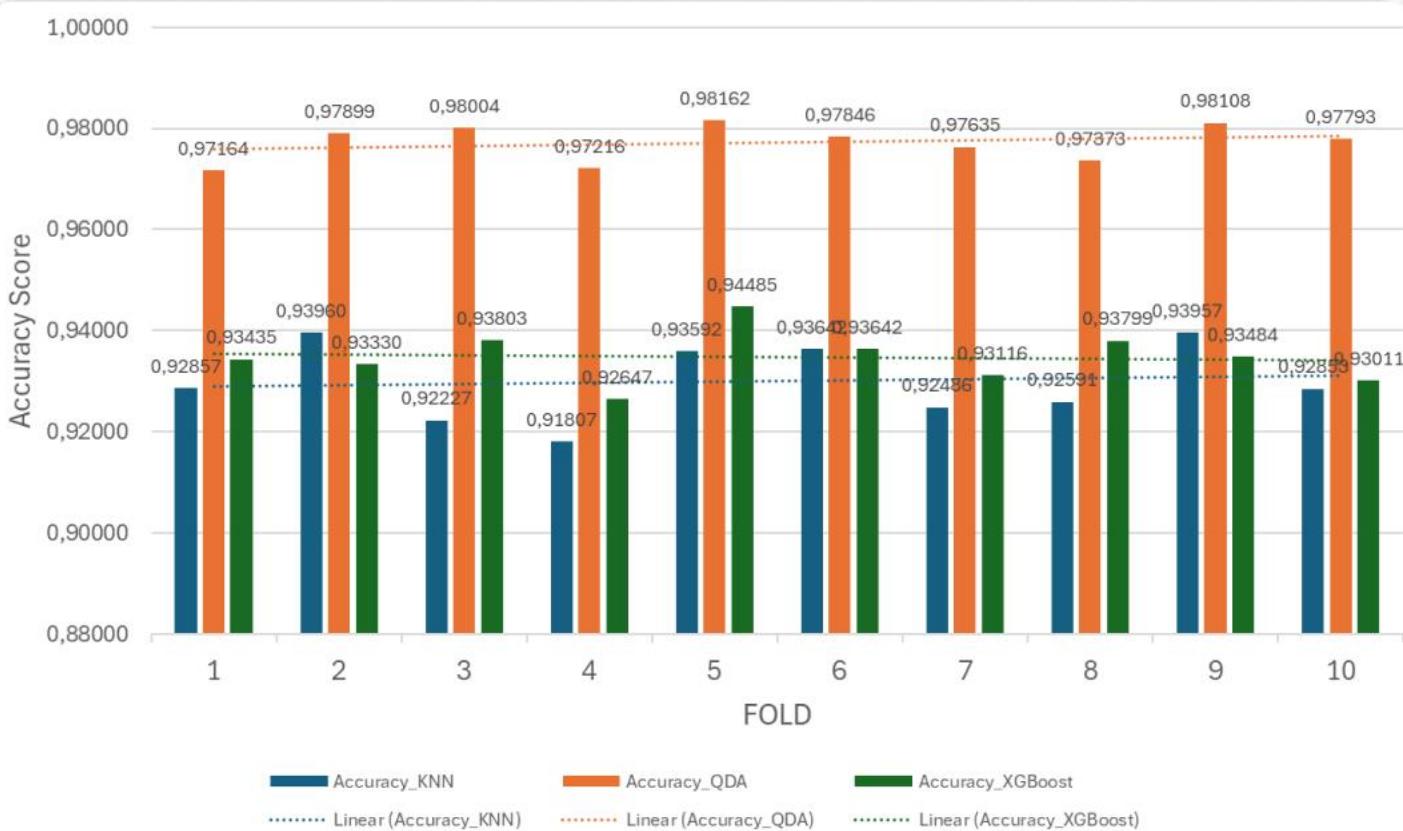
$$\text{Recall} = \frac{TP}{TP + FN}$$

----- Model Evaluation with Evaluation Metrics -----



# HASIL DAN PEMBAHASAN

## 4. Pemilihan dan Pelatihan Model



----- Metode Cross Validation Dengan Rerata Skor -----

**MEAN ± STD**

**1. ACCURACY KNN:**

**0.9300 ± 0.0071**

**2. ACCURACY QDA:**

**0.9772 ± 0.0034**

**3. ACCURACY XGBOOST:**

**0.9348 ± 0.0048**



# HASIL DAN PEMBAHASAN

## 4. Pemilihan dan Pelatihan Model

### D) PEMBUKTIAN MODEL

**Quadratic Discriminant Analysis (QDA)** mengharuskan beberapa **asumsi untuk validitas**. Asumsi yang diuji meliputi:

- 1.** *SETIAP FITUR (VARIABEL) DALAM DATASET MENGIKUTI DISTRIBUSI NORMAL UNTUK SETIAP KELAS.*
  - DISINI MENGGUNAKAN **UJI NORMALITAS DENGAN HISTOGRAM** DAN MELIHAT BAGAIMANA BENTUKNYA PADA **SETIAP FEATURE** YANG DIGUNAKAN PADA MODELING.
  
- 2.** *SETIAP KELAS MEMILIKI MATRIKS KOVARIANS YANG BERBEDA.*
  - MAKA AKAN DILAKUKAN **PENGECEKAN MENGGUNAKAN DETERMINANT MATRIKS** MASING-MASING CLASS DAN **UJI BOX'S M** UNTUK MEMERIKSA KESAMAAN KOVARIANS ANTAR KELAS

----- WHY IS QDA SUITABLE FOR THIS DATA?? LET'S PROVE IT -----



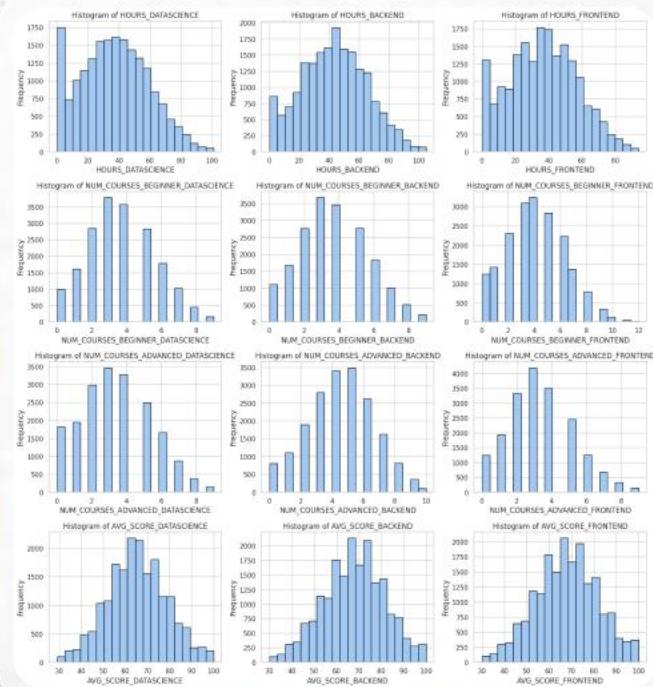
# HASIL DAN PEMBAHASAN



## 4. Pemilihan dan Pelatihan Model

1.

SETIAP FITUR (VARIABEL) DALAM DATASET MENGIKUTI DISTRIBUSI NORMAL UNTUK SETIAP KELAS.



Histogram Plot

2.

SETIAP KELAS MEMILIKI MATRIKS KOVARIANS YANG BERBEDA.

CLASS/LABEL

DETERMINAN MATRIKS KOVARIANS

1. ADVANCED\_BACKEND :

116112196910806.84

2. ADVANCED\_DATA\_SCIENCE :

104141135216291.61

3. ADVANCED\_FRONT\_END :

21081722041064.703

4. BEGINNER\_DATA\_SCIENCE :

171475279077700.12

5. BEGINNER\_FRONT\_END :

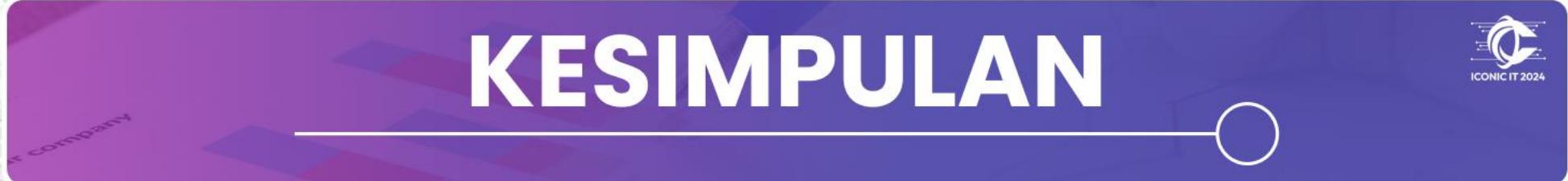
75519506539239.06

Determinan Matriks Kovarians

	Chi2	df	pval	equal_cov
box	inf	390.0	0.0	False

Uji Box's M





# KESIMPULAN

1. Penelitian ini berhasil mengembangkan model personalisasi profil pengguna berdasarkan aktivitas pembelajaran online dengan **tiga model utama: QDA, XGBoost, dan KNN**, di mana **QDA terbukti paling efektif dengan akurasi dan konsistensi tertinggi**.
2. Model QDA **mampu menangkap variasi dalam data pengguna dan menghasilkan prediksi yang akurat**, menunjukkan kekuatan model dalam mempersonalisasi profil berdasarkan aktivitas belajar.
3. **Distribusi profil pengguna merata antara level beginner dan advanced**, menandakan efektivitas metode pembelajaran untuk semua tingkat keahlian.
4. **Pengguna advanced** menghabiskan **lebih banyak waktu belajar** dan mengikuti **lebih banyak kursus** dibandingkan pengguna **beginner**.
5. **Kinerja pengguna di level advanced tidak selalu lebih tinggi**, menunjukkan tantangan dalam materi kompleks di bidang seperti data science dan frontend.



# IMPLIKASI DAN REKOMENDASI

## IMPLIKASI:

### 1. PENINGKATAN PERSONALISASI

Dengan menggunakan model QDA, platform pembelajaran dapat meningkatkan personalisasi konten yang diberikan kepada pengguna, sehingga lebih sesuai dengan kebutuhan dan kemampuan individu.

### 2. Pengembangan Kurikulum yang Lebih Adaptif

Temuan bahwa pengguna dengan profil advanced sering kali menghadapi tantangan lebih besar dalam mencapai nilai tinggi menunjukkan perlunya kurikulum yang lebih adaptif dan dukungan yang lebih kuat bagi pengguna di level ini.

### 3. Optimalisasi Pengalaman Pengguna

Menyediakan rekomendasi yang lebih tepat waktu dan relevan berdasarkan prediksi profil pengguna dapat meningkatkan keterlibatan dan motivasi pengguna dalam proses belajar.

## REKOMENDASI:

1. Implementasikan model QDA sebagai algoritma utama untuk personalisasi konten di platform pembelajaran online.
2. Lakukan penyesuaian pada materi kursus yang lebih menantang untuk pengguna di level advanced, dengan menyediakan sumber daya tambahan atau metode pengajaran alternatif.
3. Terus pantau dan analisis data pengguna untuk memperbarui dan meningkatkan model prediksi secara berkala.



# KETERBATASAN PENELITIAN DAN ARAH PENELITIAN SELANJUTNYA

## 1. PENELITIANINI MEMILIKI BEBERAPA KETERBATASAN YANG PERLU DIPERHATIKAN:

- **Keterbatasan Data:** Dataset yang digunakan mungkin tidak mencakup semua variabel yang mempengaruhi kinerja pembelajaran pengguna. Selain itu, data yang digunakan terbatas pada platform tertentu, sehingga hasilnya mungkin tidak sepenuhnya generalizable.
- **Asumsi Model:** Model QDA mengasumsikan distribusi normal dari data, yang mungkin tidak sepenuhnya tercermin dalam dataset aktual. Meskipun demikian, performa model tetap cukup baik dalam konteks penelitian ini.

## 2. UNTUK PENELITIAN SELANJUTNYA, BEBERAPA LANGKAH DAPAT DIAMBIL:

- **Pengujian Model di Platform Lain:** Menerapkan dan menguji model di platform pembelajaran online lainnya untuk melihat apakah hasil yang serupa dapat diperoleh.
- **Penggunaan Dataset yang Lebih Luas dan Beragam:** Melibatkan dataset dari berbagai sumber dan jenis aktivitas pembelajaran yang lebih beragam untuk meningkatkan generalisasi model.
- **Eksplorasi Model Lain:** Menguji model machine learning lainnya yang mungkin lebih cocok untuk jenis data yang berbeda atau untuk menyelesaikan masalah lain dalam konteks pembelajaran online.





Present by: ROYALS OUTLAWS\_

# THANK YOU FOR YOUR ATTENTION



ANY  
QUESTION???

