

# STUDENT'S PERFORMANCE PREDICTION

Tugas AOL MLOps

# ANGGOTA KELOMPOK

- Brandon Tumiwa - 2802508224
- Justine Ria Jingga - 2802536783
- Stanley Angkasa - 2802550642
- Raka Priyahita Pramudito - 2802544500


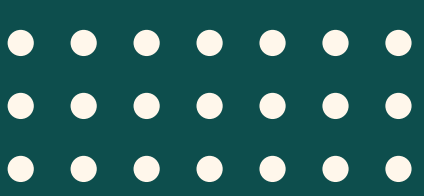
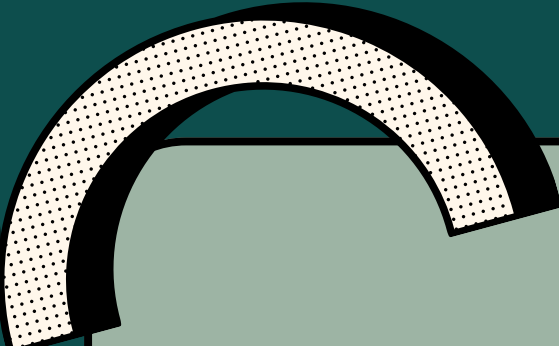


# LATAR BELAKANG

- Pendidikan dipengaruhi banyak faktor akademik & non-akademik
- Machine Learning dapat mempelajari pola performa siswa
- Tujuan: memprediksi nilai ujian akhir siswa

# DESKRIPSI DATASET

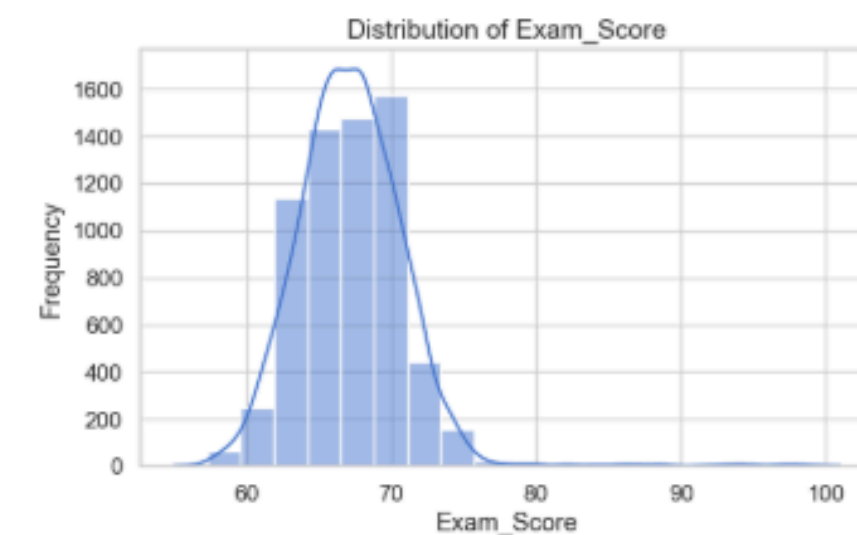
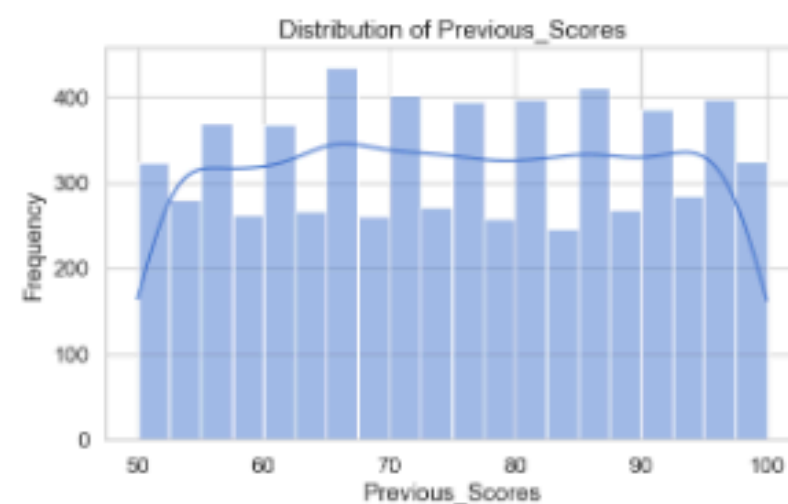
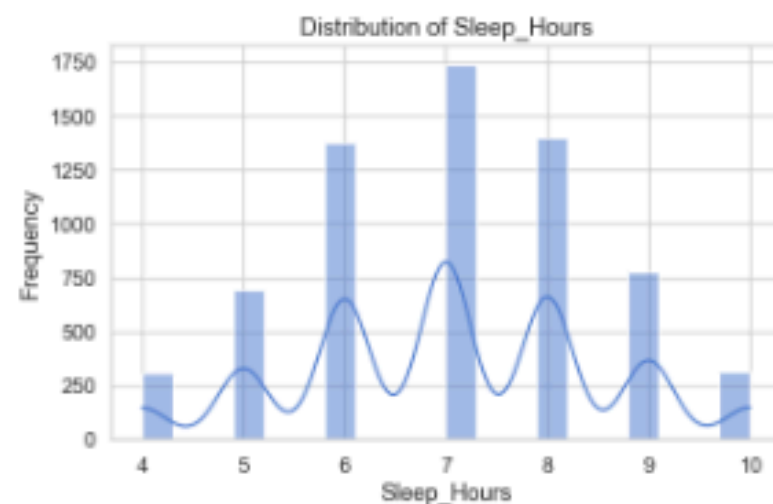
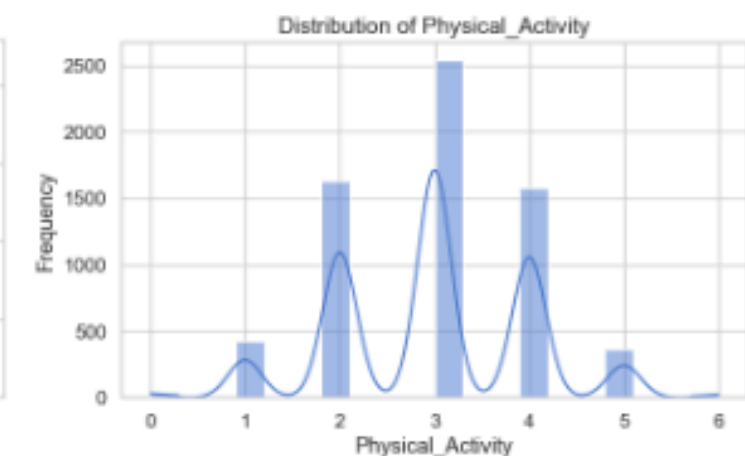
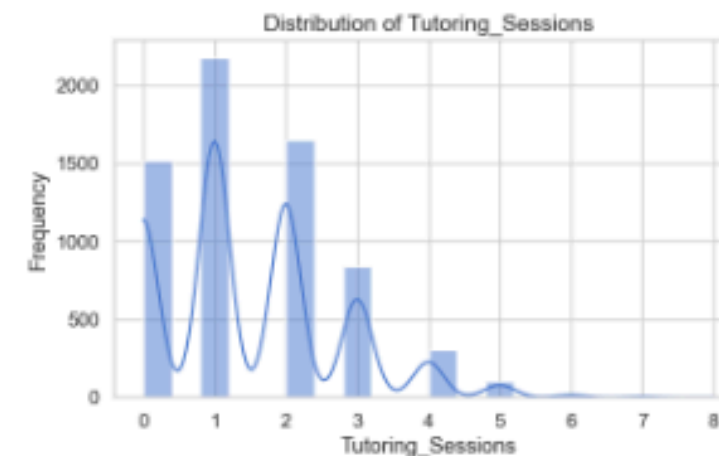
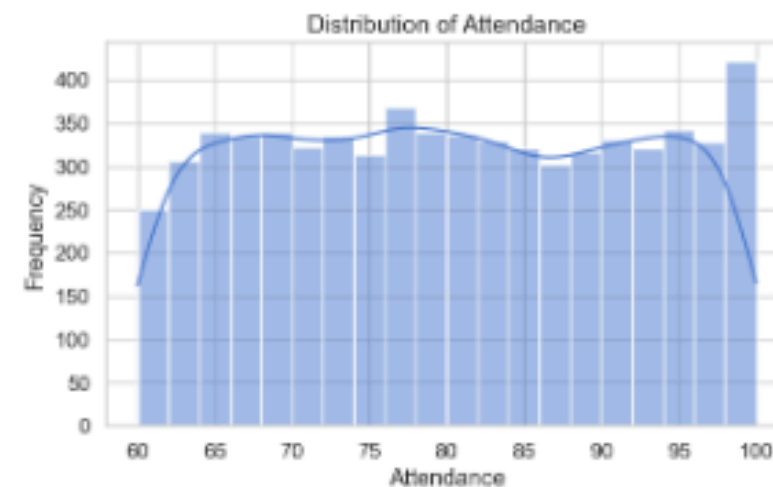
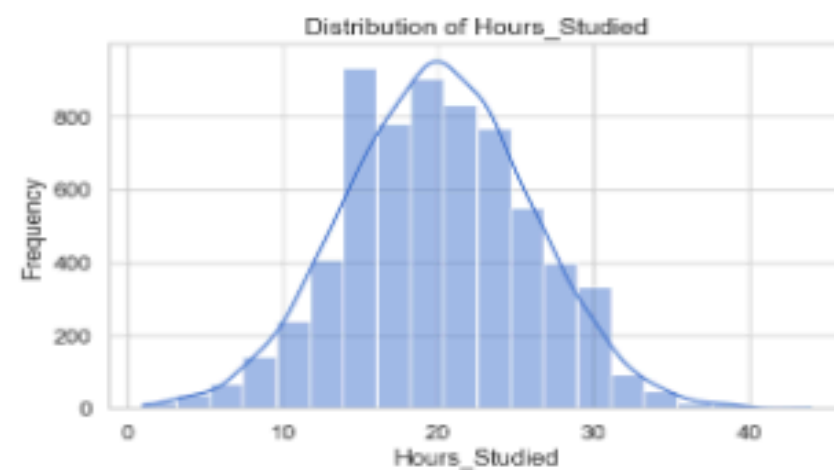
- Dataset: Student Performance Factors (Kaggle)
- Total data: ±6.600 siswa
- 19 fitur input + 1 target (Exam\_Score)
- Fitur numerik & kategorikal
- Mencakup faktor akademik, keluarga, dan lingkungan



# EXPLORATORY DATA ANALYSIS (EDA)

# ANALISIS FITUR NUMERIK

- tipe data konsisten, range nilai wajar ( min 55 max 101, mean 67,2 )
- Analisis distribusi (univariate) menggunakan histogram



# ANALISIS FITUR NUMERIK

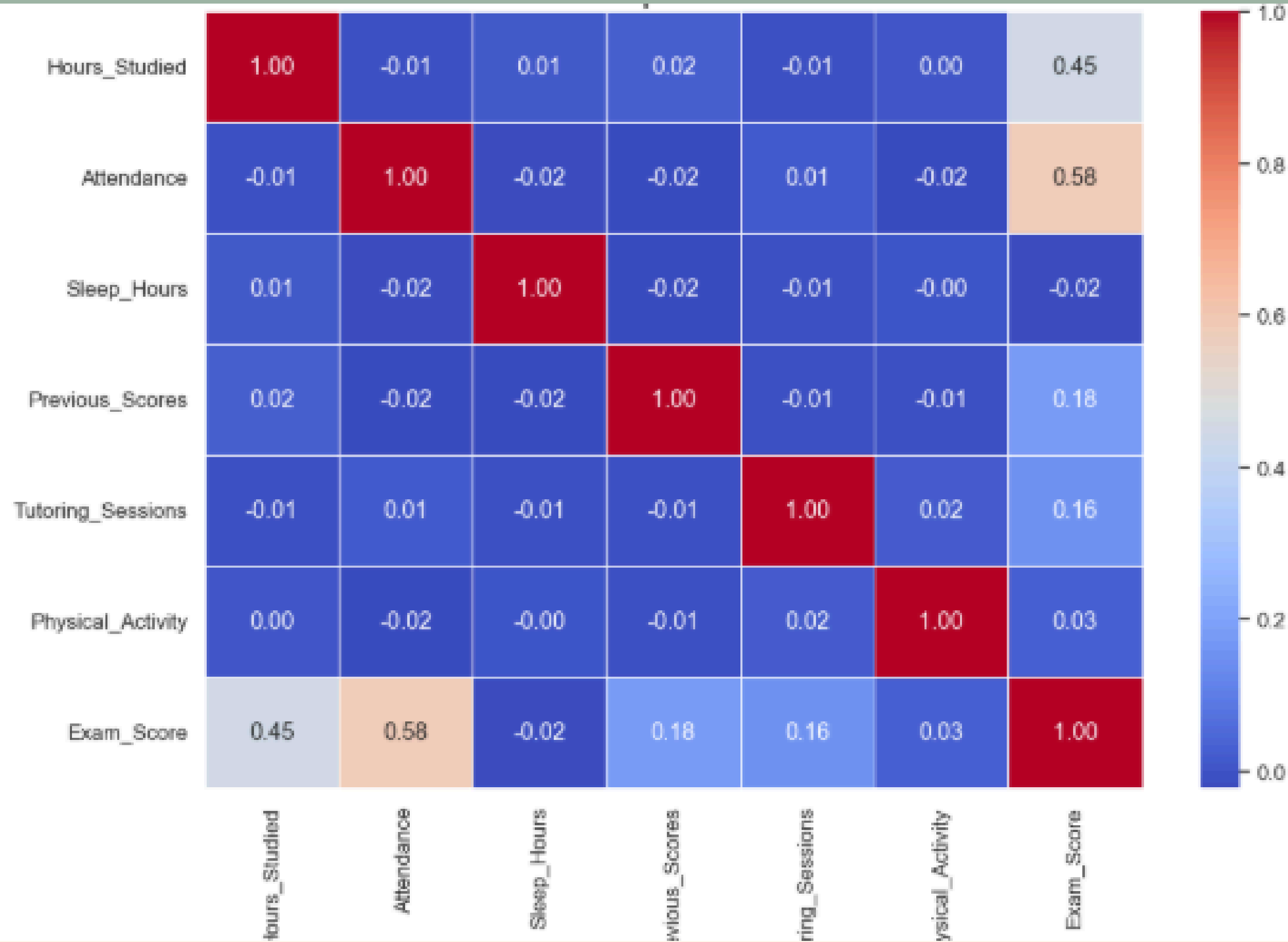
- ANALISIS OUTLIER

Menggunakan visualisasi Boxplot

Jumlah outlier yang didapatkan dari perhitungan :

- Hours\_Studied: 43 rows
- Tutoring\_Sessions : 430 rows
- Exam Scores : 104 rows
- Attendance, Sleep\_Hours, Previous\_Scores, Physical\_Activities: tidak ada outlier

# ANALISIS KORELASI





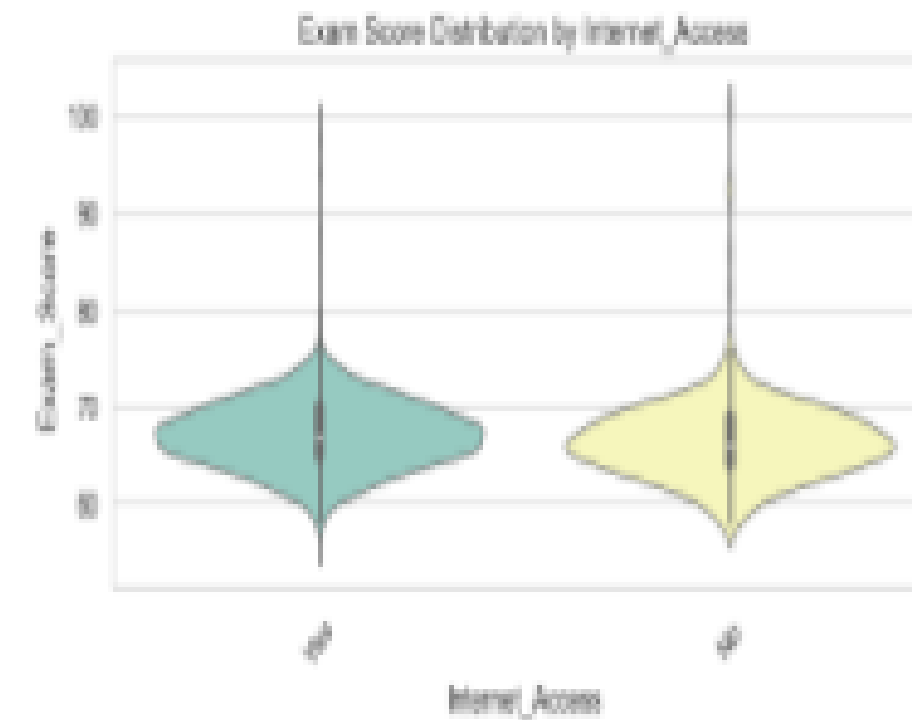
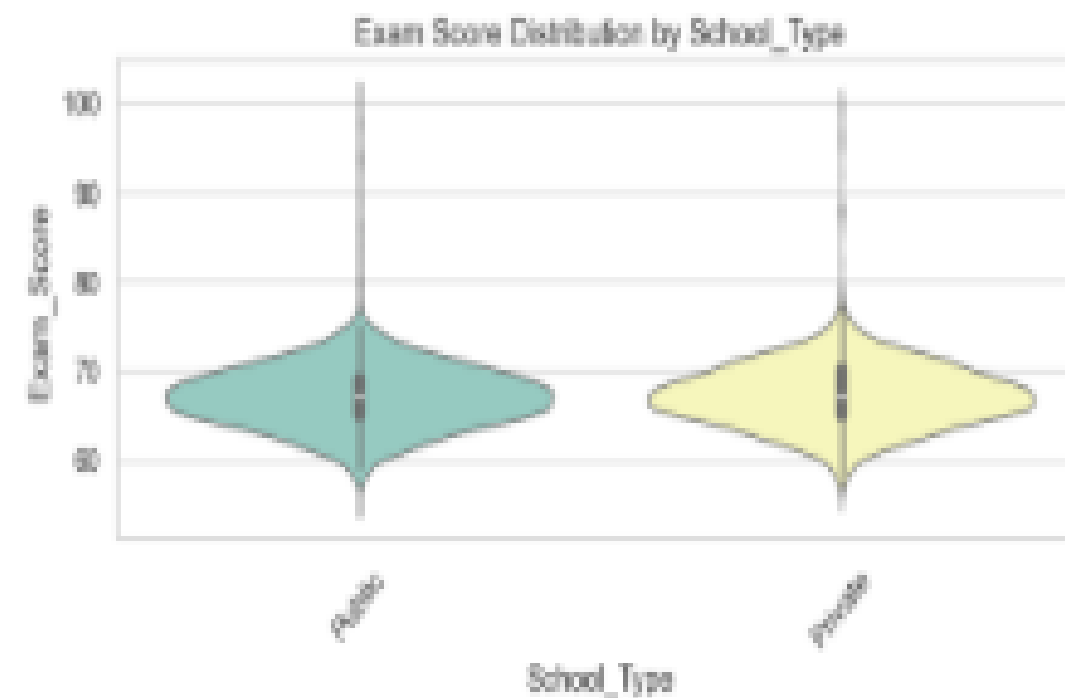
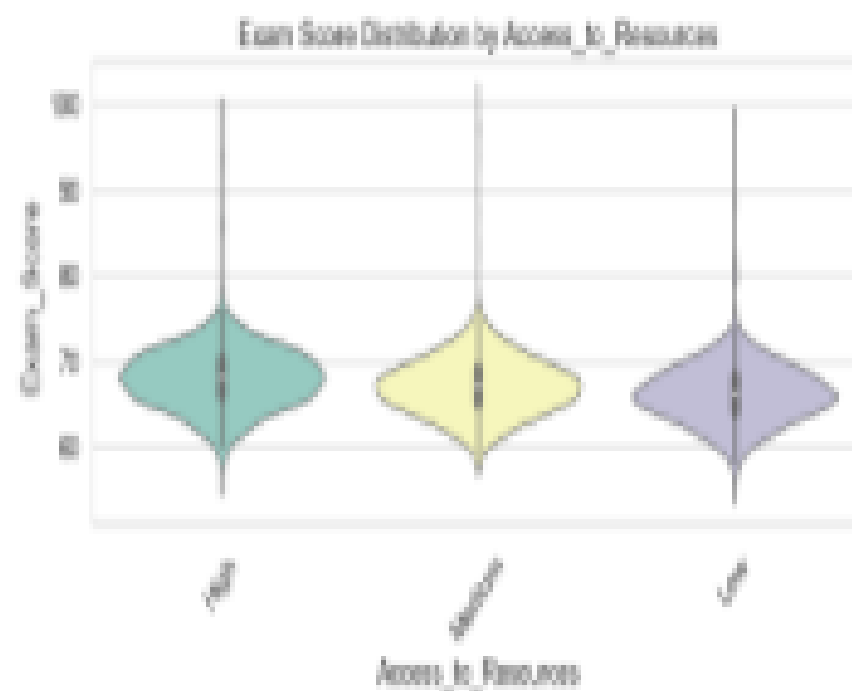
# ANALISIS FITUR KATEGORIKAL

- Analisis distribusi fitur (univariate) menggunakan Countplot
- Ada fitur yang cenderung seimbang
- Ada yang relatif seimbang, dengan fitur tengah mendominasi
- Ada yang menunjukkan ketidakseimbangan yang cukup besar, di mana terdapat kategori mayoritas yang sangat dominan

Hasil ini memberikan informasi penting bagi tahap preprocessing, terutama dalam pemilihan metode encoding yang sesuai.

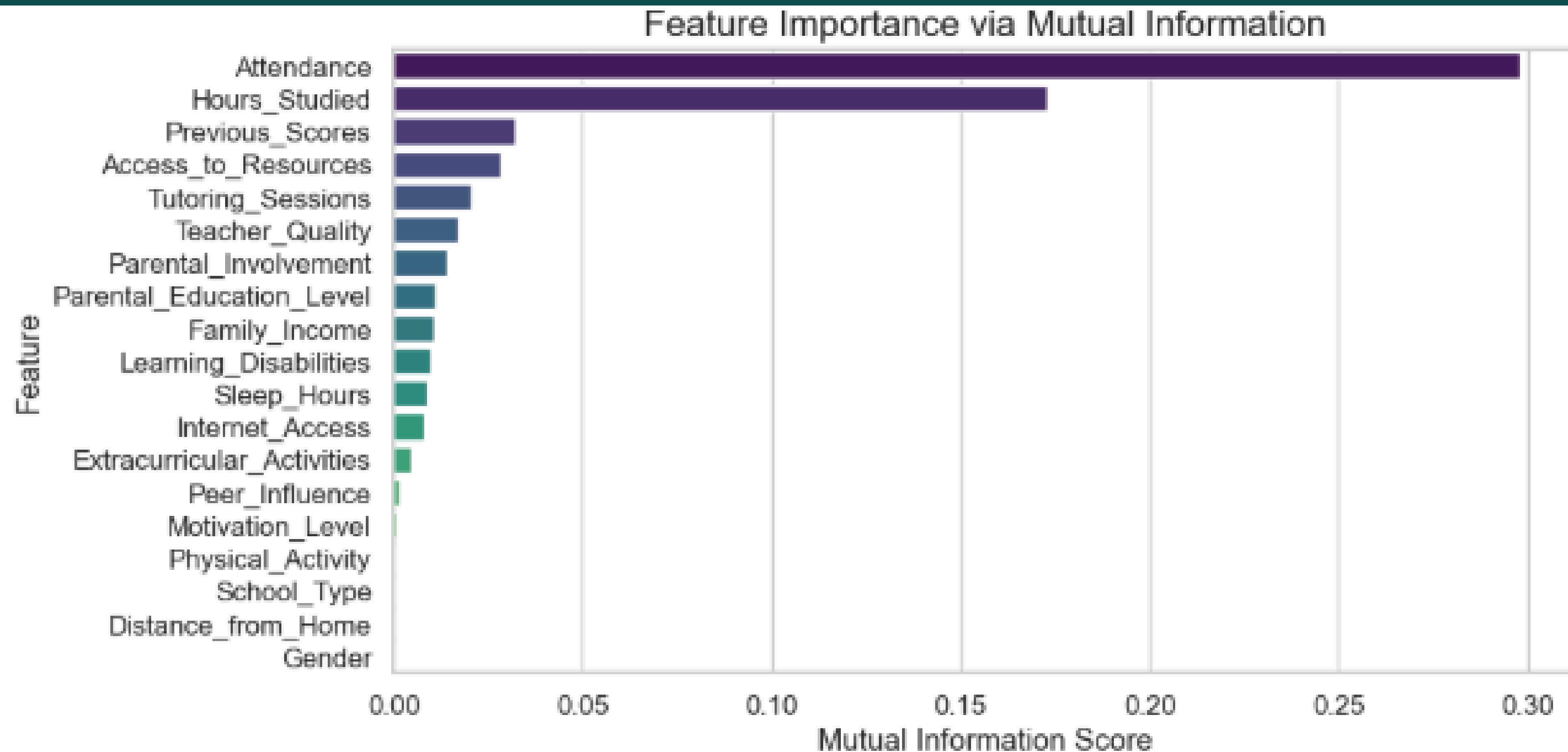
# ANALISIS FITUR KATEGORIKAL

- Analisis fitur kategorikal terhadap target menggunakan Violin Plot  
Sebagian besar kategori menunjukkan median nilai ujian yang berada pada rentang yang sama, dengan sebaran distribusi yang relatif seragam.



# MUTUAL INFORMATION

Mutual Information (MI) digunakan untuk mengukur hubungan non-linear antara fitur input dan variabel target Exam\_Score.



# PREPROCESSING

# PREPROCESSING

- Feature Grouping (numerik, binary, ordinal, nominal), memastikan setiap fitur diproses secara tepat

## Missing Values Handling

- Fitur numerik → diisi dengan median
- Fitur kategorikal → diisi dengan modus (most frequent)

Data dibagi menjadi 80% data training dan 20% testing

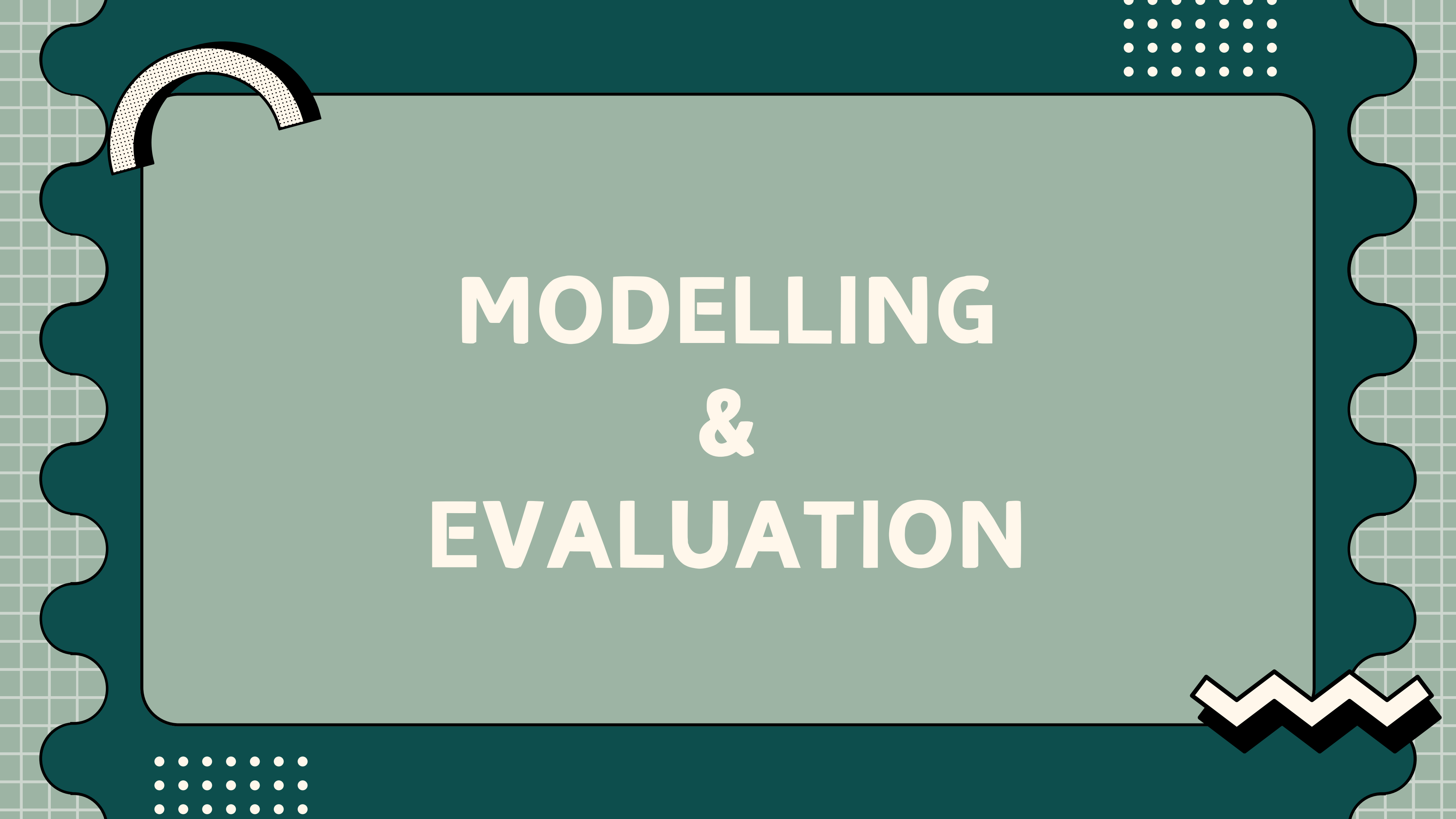
# PREPROCESSING

## Encoding Strategy

- Binary encoding (0/1) untuk fitur dua kelas
- Ordinal encoding untuk fitur berurutan (Low → Medium → High)
- One-hot encoding untuk fitur nominal tanpa urutan

Fitur numerik diproses dengan StandardScaler

Seluruh preprocessing digabung dalam Pipeline & ColumnTransformer



# MODELLING & EVALUATION

# MODEL

- Model yang digunakan : XGBoost Regressor

Alasan pemilihan XGBoost:

- Sangat kuat untuk data tabular
- Mampu menangkap hubungan non-linear
- Robust terhadap outlier
- Memiliki regularisasi bawaan

Model digabung dengan preprocessing menggunakan Pipeline (dipakai untuk tuning)



# HYPERPARAMETER TUNING

**Metode tuning :** RandomizedSearchCV

**Cross-validation :** 3-fold

**Metric optimasi :** MAE (Mean Absolute Error)

**Parameter yang dituning:** n\_estimators, learning\_rate, max\_depth, subsample, colsample\_bytree, gamma, min\_child\_weight

Tuning membantu menemukan kombinasi parameter terbaik tanpa overfitting.


# EVALUASI

Setelah tuning, model diuji pada data testing :

- MAE sekitar 0.59
- RMSE sekitar 1.86

Error sangat rendah karena variasi target (nilai ujian) di dataset juga kecil

- CV MAE  $\approx 0.65$ , Test MAE  $\approx 0.59$ , maka model generalize dengan baik.
- $R^2$  sekitar 0.76, model berhasil menjelaskan sekitar 76% variasi data



# DEPLOYMENT

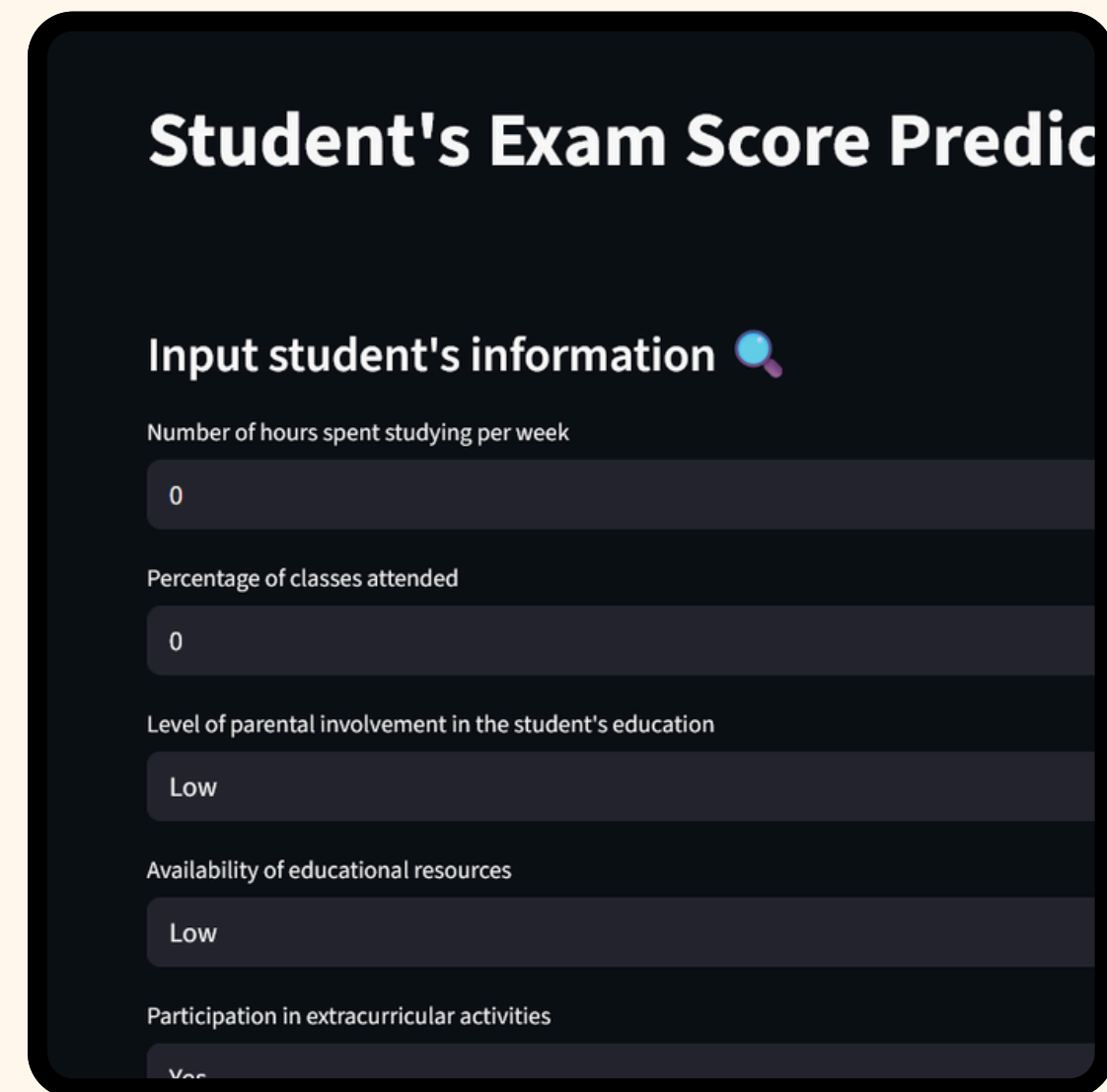
# DEPLOYMENT APLIKASI

- Model prediksi nilai ujian berbasis XGBoost diintegrasikan ke dalam aplikasi web menggunakan Streamlit
- Aplikasi kemudian di-deploy ke Streamlit Cloud



# IMPLEMENTASI & INTEGRASI MODEL

- Aplikasi dikembangkan menggunakan framework Streamlit karena sederhana dan interaktif
- Fitur interface yang disediakan:
  - Input numerik
  - Input kategorikal dan biner
- Model XGBoost yang telah dilatih disimpan dalam bentuk file .pkl
- Model dan file fitur dimuat ke aplikasi tanpa pelatihan ulang
- Data input diproses dan dikonversi sesuai struktur model sebelum menghasilkan prediksi



The screenshot shows a web application titled "Student's Exam Score Prediction". Below the title is a section labeled "Input student's information" with a magnifying glass icon. There are five input fields, each with a label and a value:

- Number of hours spent studying per week: 0
- Percentage of classes attended: 0
- Level of parental involvement in the student's education: Low
- Availability of educational resources: Low
- Participation in extracurricular activities: Yes



# PROSES DEPLOYMENT & PENGUJIAN

- Repository GitHub dihubungkan ke Streamlit Cloud
- File utama aplikasi (app.py) dijadikan entry point
- Streamlit Cloud secara otomatis:
  - Menginstal dependensi dari requirements.txt
  - Menjalankan aplikasi di lingkungan cloud
  - Aplikasi dapat diakses melalui URL publik

# APLIKASI

link dataset : <https://studentperformance-kkynpbuxtwx8ejdfysnyab.streamlit.app/>



# TERIMA KASIH

Disusun Oleh:  
Brandon Tumiwa, Justine Ria Jingga, Raka Priyahita Pramudito, Stanley Angkasa