

TUGAS UTS MACHINE LEARNING

ARDIUS EBENEZER SIMANJUNTAK

1103210208

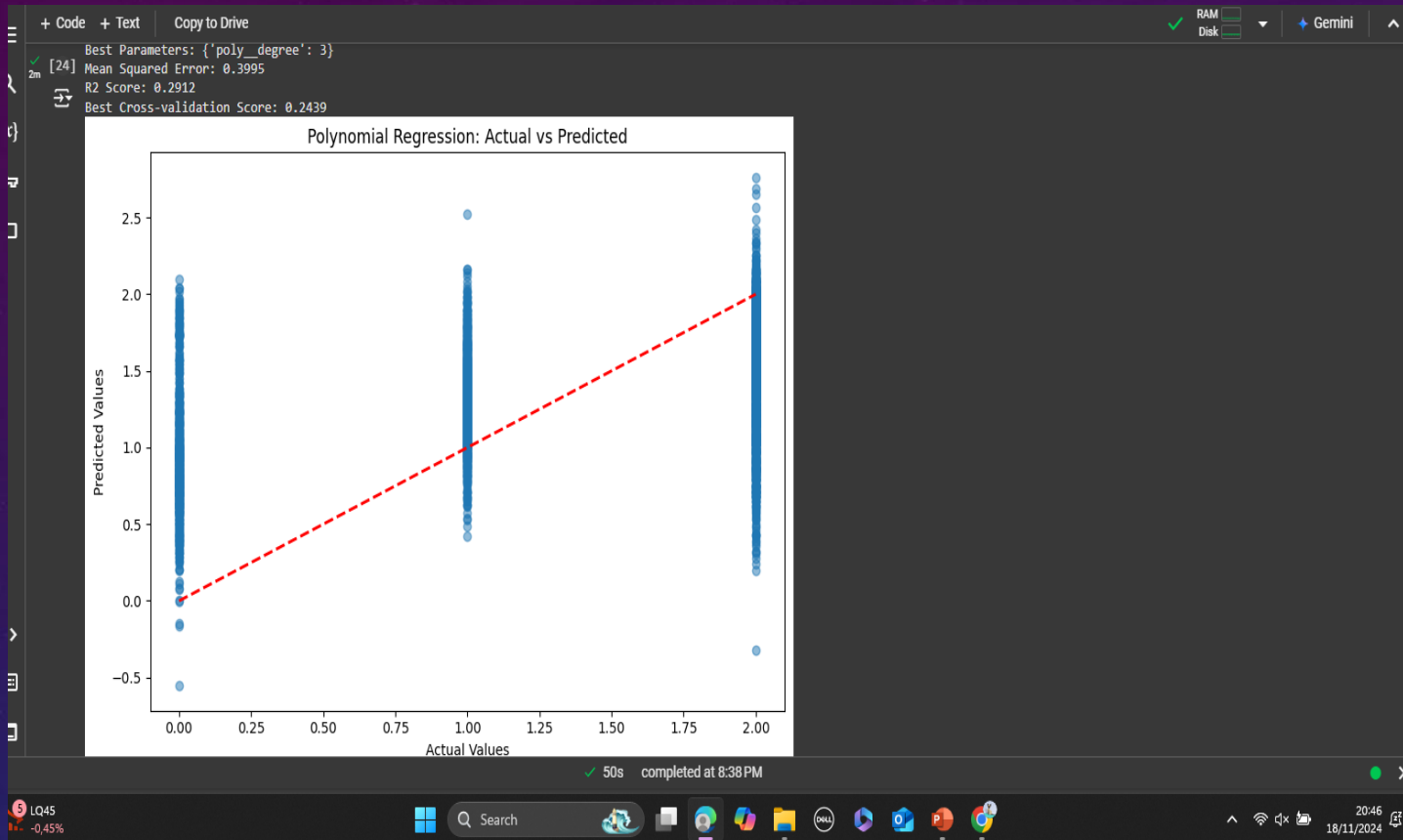
EXPLORATORY DATA ANALYSIS

EXPLORATORY DATA ANALYSIS (EDA) MERUPAKAN PROSES BAGIAN YANG PENTING SEBELUM MELAKUKAN FEATURE ENGINEERING DAN MODELING. PROSES INI MENGANALISIS DATA AWAL UNTUK MEMAHAMI STRUKTUR, KOMPONEN PENTING DAN KARAKTERISTIK DATA SEBELUM PEMODELAN DATA. TUJUAN DALAM MODEL INI UNTUK MENDAPATKAN GAMBARAN UMUM UNTUK BISA MELAKUKAN ANALIS TERHADAP KUMPULAN DATA SECARA MENDALAM.

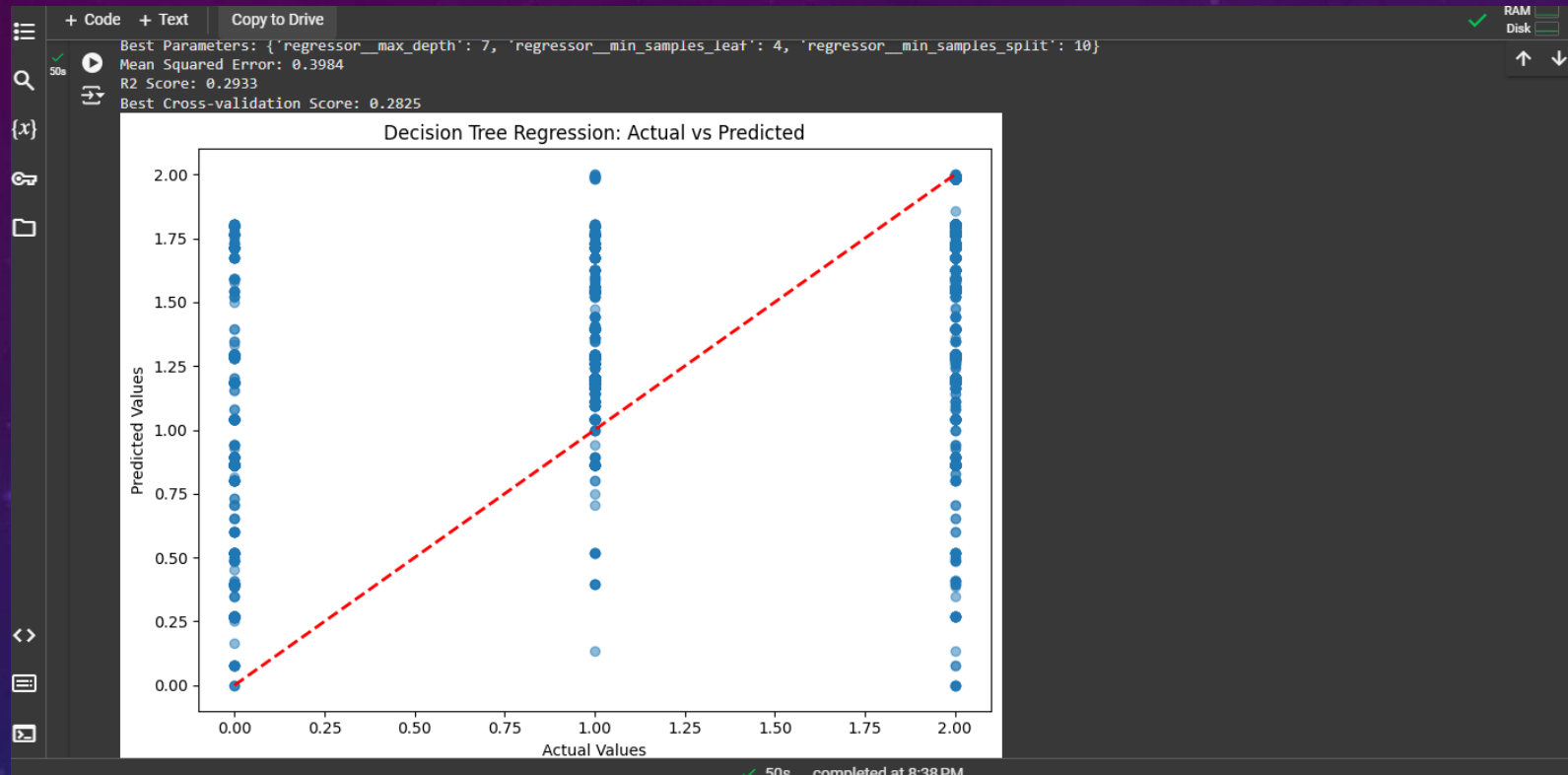
VISUALISASI DATA MUNCUL DALAM MEMBUAT REPRESENTASI VISUAL DALAM MEMBANTU PENGGUNA DALAM MEMAHAMI POLA, TREND, DIAGRAM YANG TERSEMBUNYI DI DALAM DATA. INI MENJADI KUNCI DALAM EDA UNTUK MEMBUAT PROSES ANALISIS DATA SECARA EKSPLORATIF YANG MENJADI LEBIH EFISIEN.

CLASSIFICATION MODEL

The background is a gradient of dark blue and purple, speckled with white dots resembling stars. On the right side, there are faint, light-blue technical diagrams. These include a large circular scale with degree markings from 0 to 210, a smaller concentric circle with an arrow, and a dashed circular path with an arrow. In the bottom left corner, there is another dashed circular path with an arrow.

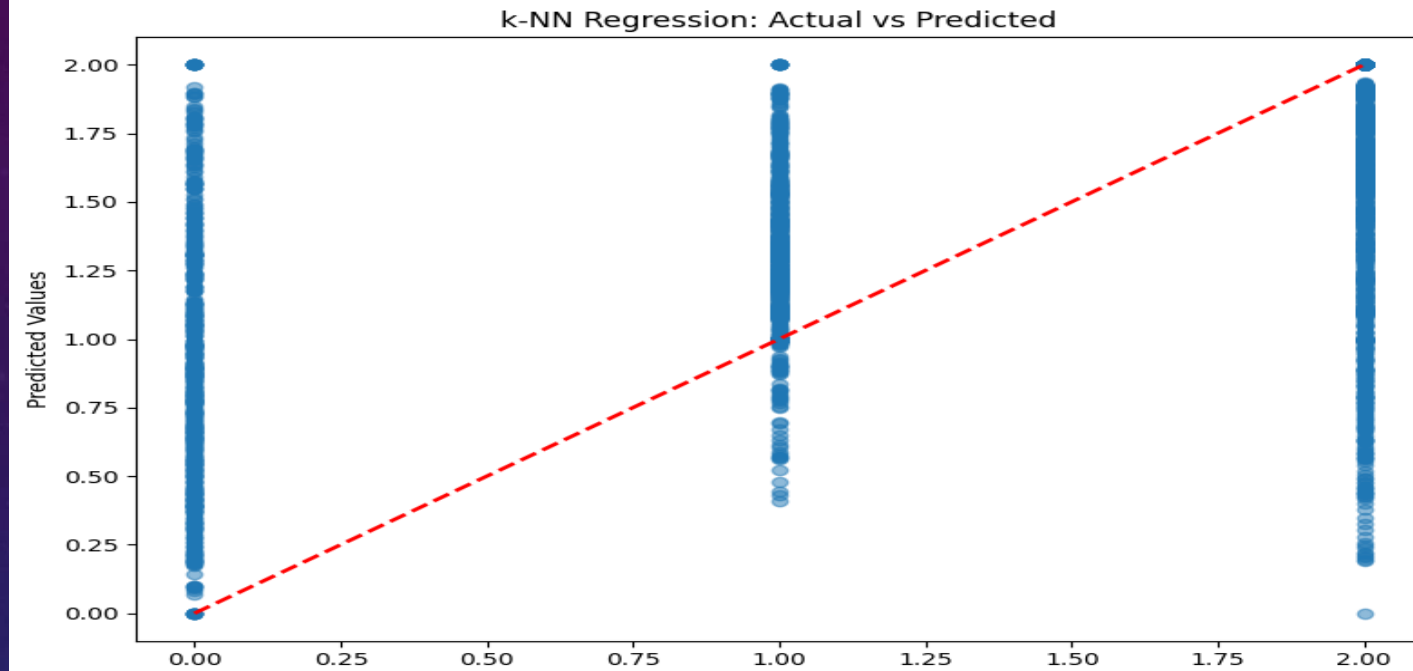


Logistic Model, Model ini menampilkan dengan metode statistik yang mengklasifikasi kelas biner dengan cara memodelkan probabilitas keluaran menggunakan fungsi sigmoid. Model yang dihasilkan bersifat koefisien linear, Probabilitas dan Prediksi Akhir dari Variabel hasil yang bersifat dikotomis, yang berarti hanya ada dua kelas yang mungkin.



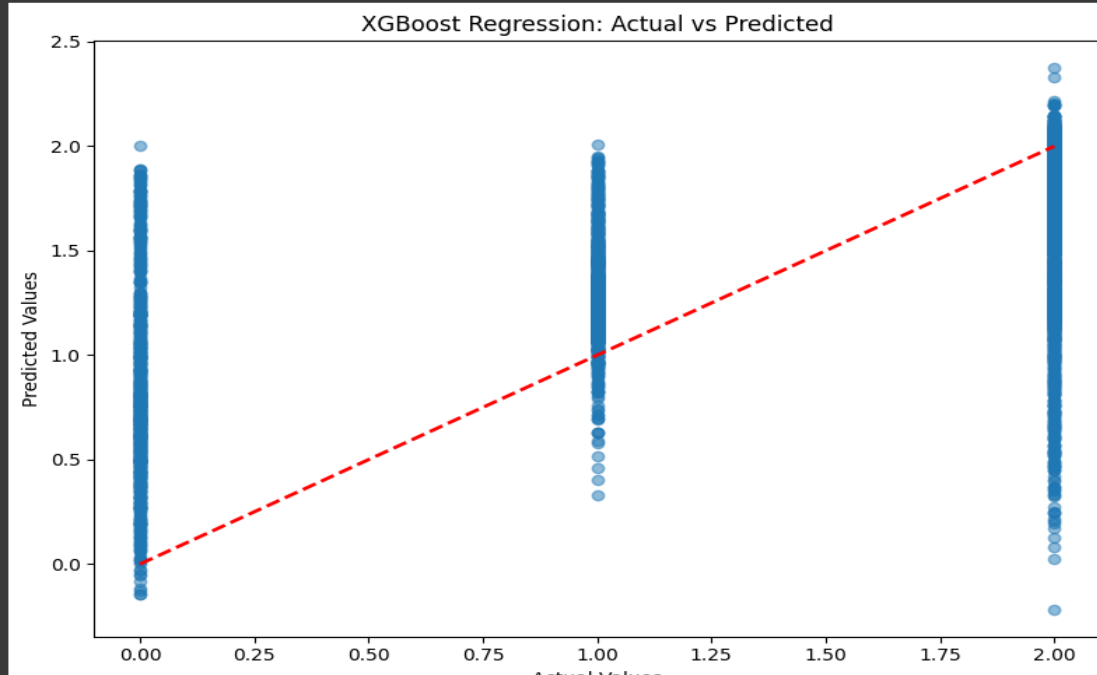
Decision Tree, Model ini dihasilkan dengan pengambilan Keputusan berbasis serangkaian aturan if-else dengan prediksi output yang berdasarkan data set yang diberikan. Biasanya representasi visual decision tree ini berbentuk diagram pohon.

```
Best Parameters: {'regressor__n_neighbors': 9, 'regressor__p': 1, 'regressor__weights': 'distance'}  
Mean Squared Error: 0.3607  
R2 Score: 0.3601  
Best Cross-validation Score: 0.3077
```



k-NN (k-Nearest Neighbors) merupakan model yang ada pada pustaka Scikit-learn sebagai metode klasifikasi yang menentukan label kelas suatu data berdasarkan label kelas dari data yang paling dekat dengannya dalam ruang fitur.

```
Best Parameters: {'regressor__learning_rate': 0.1, 'regressor__max_depth': 7, 'regressor__n_estimators': 200, 'regressor__subsample': 1.0}
Mean Squared Error: 0.3446
R2 Score: 0.3887
Best Cross-validation Score: 0.3642
```



XGBoost (Extreme Gradient Boosting) adalah algoritma pembelajaran mesin berbasis pohon (tree-based) yang dirancang untuk menghasilkan model prediksi yang sangat akurat. Untuk tugas Classification, model XGBoost menghasilkan rangkaian pohon keputusan (decision trees) yang digabungkan secara iteratif melalui pendekatan gradient boosting.