**DataMining And WareHousing**

**Name:Ali Naqi**

**Seat#: B18101011**

**Data Preprocessing**

|  |  |
| --- | --- |
| **source** | **https://www.kaggle.com/datasets/mohamedhanyyy/top-football-leagues-scorers** |
| **Colum count** | **11** |
| **Missing values** | **none** |
| **Outliers** | **none** |
| **Normalization** |  |
| **Training Data**  **Testing Data** | **3529**  **7163** |
| **Model Evalvation:**  **1-Navie Bayes**  accuracy 0.00 3529  macro avg 0.00 0.00 0.00 3529  weighted avg 0.00 0.00 0.00 3529  Confusion Matrix:  [[0 0 0 ... 0 0 0]  [0 0 0 ... 0 0 0]  [0 0 0 ... 0 0 0]  ...  [0 0 0 ... 0 0 0]  [0 0 0 ... 0 0 0]  [0 0 0 ... 0 0 0]]  Accuracy of Decision Tree with Gini Index : 0.0019835647492207425  **2-Logistic Regression**  **3-Decision Tree –Enotropy**  **4-Decision Tree \_Gini Index** | accuracy 0.00 3529.0  macro avg 0.00 0.00 0.00 3529.0  weighted avg 0.00 0.00 0.00 3529.0  Confusion Matrix:  Accuracy of Naive Bayes : 0.0  accuracy 0.01 3529  macro avg 0.00 0.00 0.00 3529  weighted avg 0.00 0.01 0.00 3529  Accuracy of Logistic Regression : 0.010201190138849533  accuracy 0.01 3529  macro avg 0.00 0.00 0.00 3529  weighted avg 0.00 0.01 0.00 3529  Confusion Matrix:    Accuracy of Decision Tree with Entropy : 0.0076508926041371494  accuracy 0.00 3529  macro avg 0.00 0.00 0.00 3529  weighted avg 0.00 0.00 0.00 3529  Confusion Matrix:    Accuracy of Decision Tree with Gini Index : 0.0019835647492207425 |