**Character Recognition in Natural Images :**

**Description Dataset (Chars74k):**

The Chars74k dataset is a collection of images designed for training and evaluating character recognition models. It contains over 74,000 images of characters, including both handwritten and computer-rendered text.

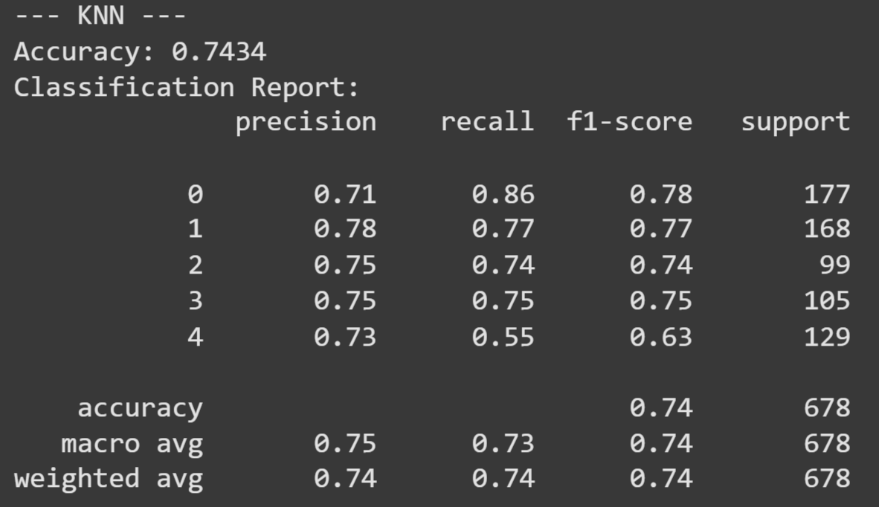
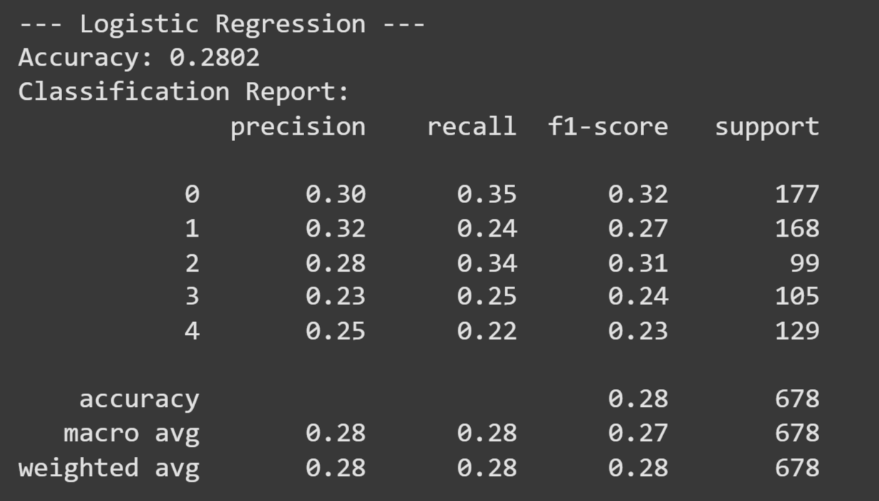
* **Type**: Contains labeled images for classification tasks.
* **Classes**: Includes 5 distinct classes: ['A', 'E', 'R', 'O', 'N'].
* **Labels**: Represent digits corresponding to the class of each image.
* **Missing Data**: None identified.
* **Size of Each Image:** Updated to 28x28 pixels (normalized and resized).
* **Training/Testing Split:** 80% training ,20% testing .

**Algorithms:**

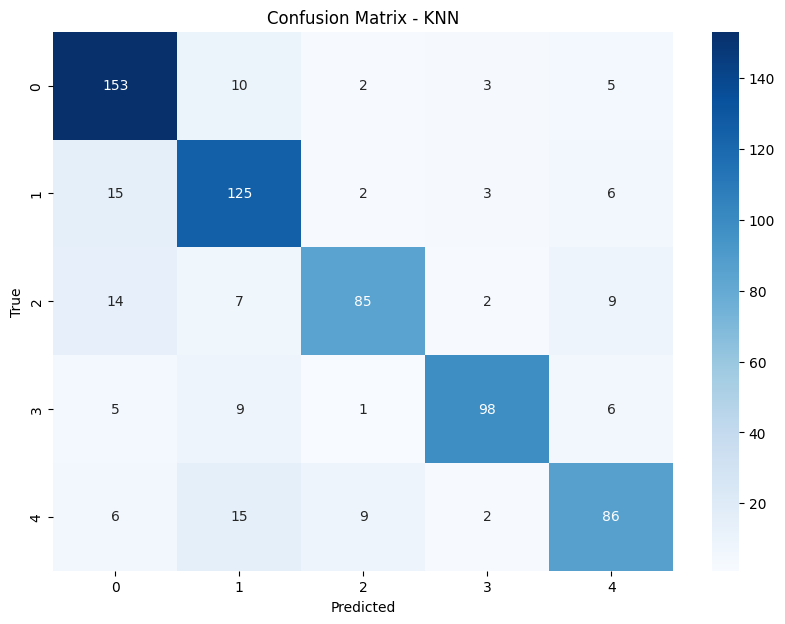
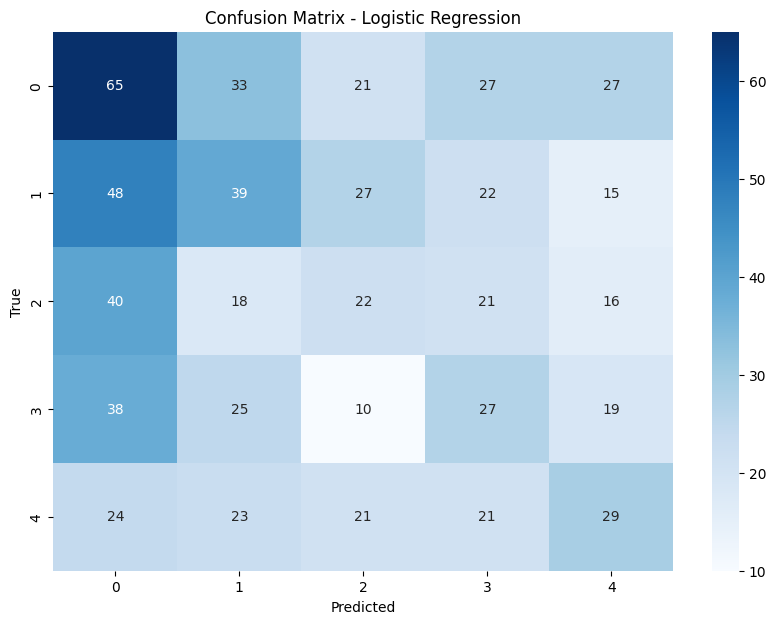
* **Logistic Regression:**
* Logistic regression is a linear model that estimates the probability of a class label using the logistic (sigmoid) function. It assumes a linear relationship between the input features and the log-odds of the target classes. While it is computationally efficient and interpretable, it struggles with non-linear or complex relationships.
* **Strengths:** Fast to train, simple to implement, interpretable results.
* **Weaknesses:** Poor performance on non-linearly separable data or complex datasets.
* **K-Nearest Neighbors (KNN):**
* KNN is a non-parametric, instance-based algorithm. It classifies a data point based on the class labels of its nearest neighbors (measured using distance metrics such as Euclidean). The majority class among the nearest neighbors determines the prediction.
* **Strengths:** Effective on non-linear data, no prior assumptions about data distribution.
* **Weaknesses:** Computationally expensive for large datasets, sensitive to irrelevant features.

**Comparison:**

* K-Nearest Neighbors (KNN) outperforms Logistic Regression on the Chars74k dataset. While Logistic Regression relies on linear decision boundaries and struggles with non-linear data (evidenced by low precision, recall, and AUC), KNN handles non-linear relationships effectively, achieving higher accuracy, precision, recall, and F1-score. Despite its higher computational cost, KNN's flexibility makes it the better choice for this classification task
* **Evaluation metrics:**

****

* **Confusion Matrix**:



* **ROC and AUC Curve:**

