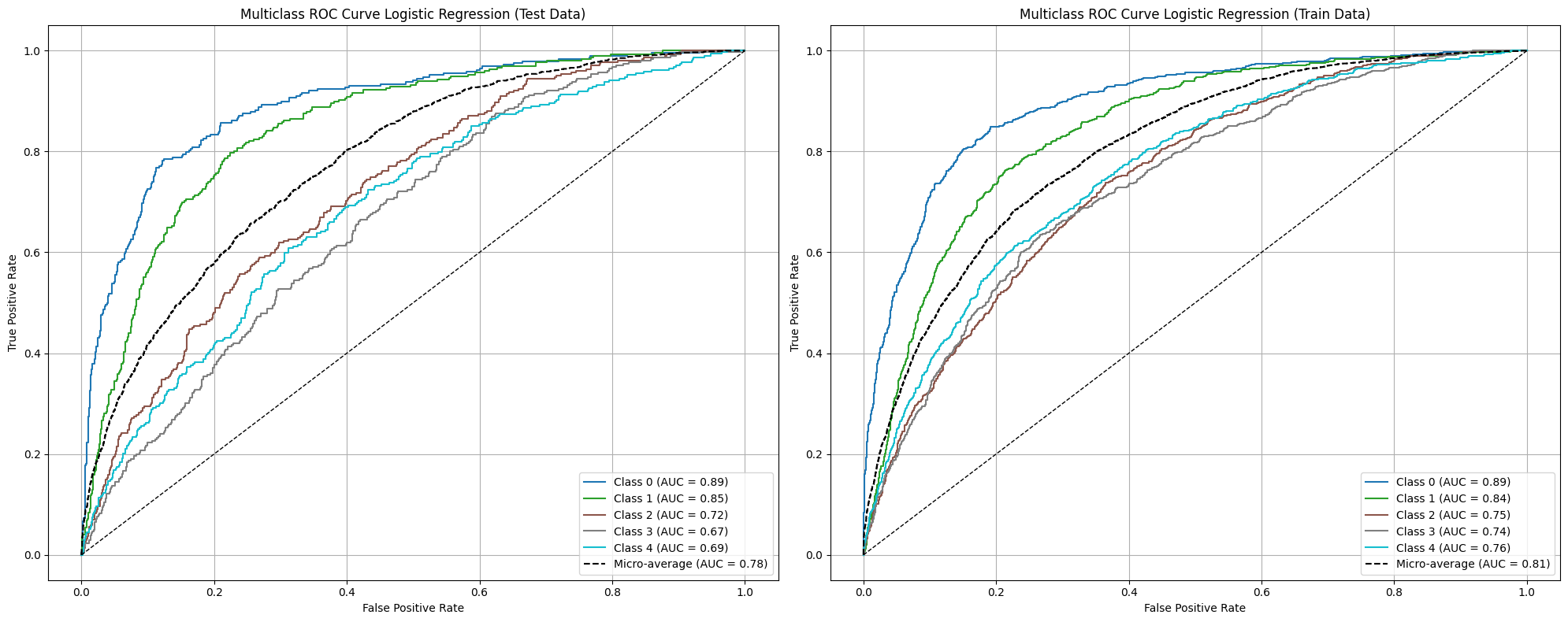
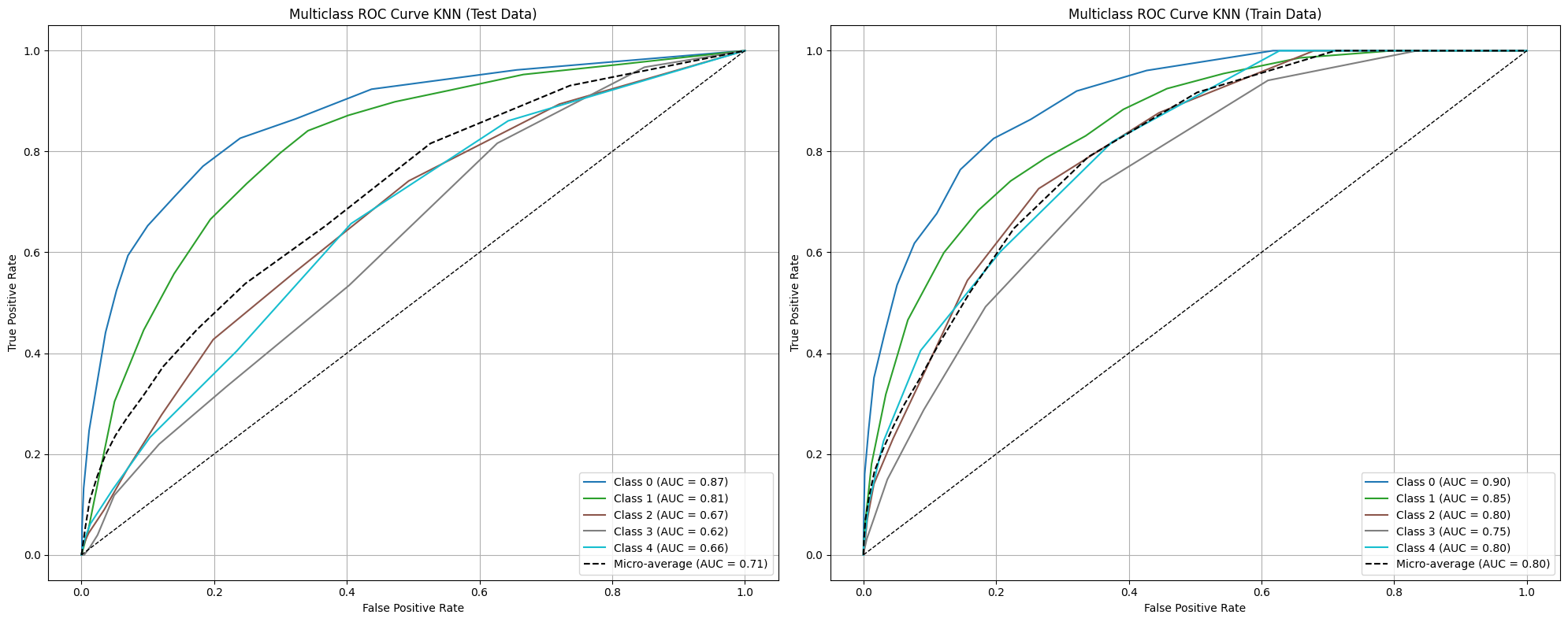
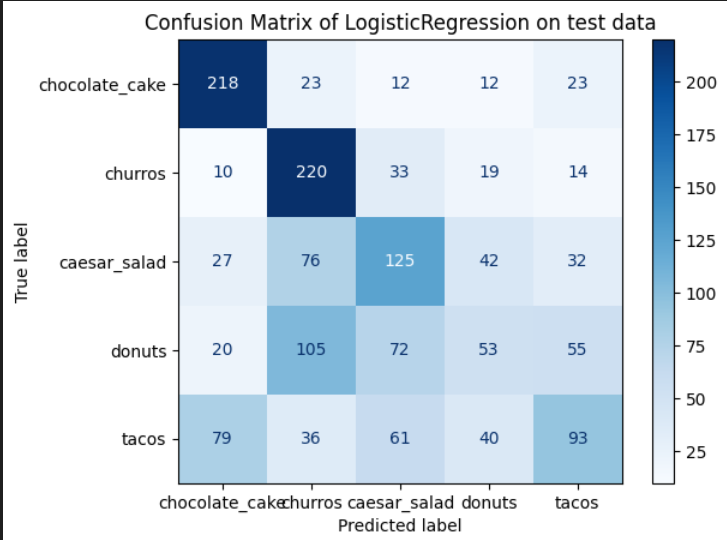
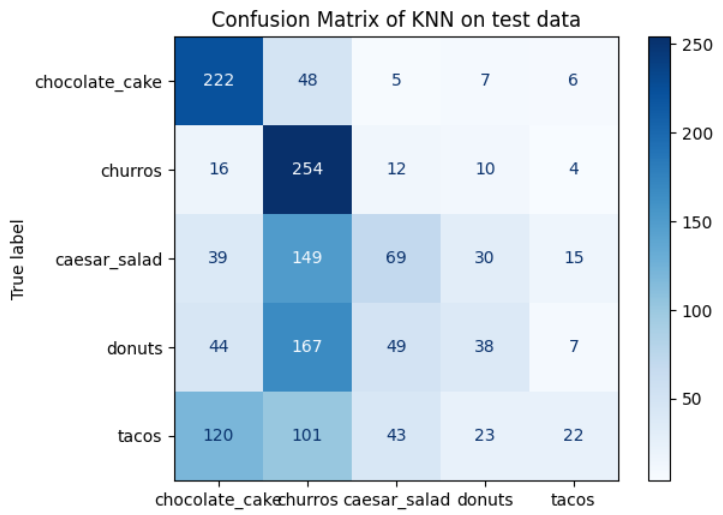
* Classification:
  + Dataset Description:
    - 5 classes each 1000 photo (Caesar\_salad, chocolate\_cake, churros, donuts, tacos)
    - Image size (512 \* 384)
    - Total size of data : 5000 image
    - Train data: 3500 Image
    - Test data : 1500 image
  + Preprocessing:
    - Resizing the images to 64 \* 64
    - Reading images with lab color space
    - Normalization each pixel by dividing it by 255 so each value be between 0 and 1
    - PCA
  + Models comparison:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Metric | Logistic regression (train) | Logistic regression (test) | KNN (train) | KNN (test) |  |  |  |  |  |  |  |
| accuracy | 50% | 47% | 44% | 40% |  |  |  |  |  |  |  |
| ROC\_AUC | 0.79 | 0.76 | 0.82 | 0.72 |  |  |  |  |  |  |  |
| precision | 49% | 45% | 49% | 40% |  |  |  |  |  |  |  |
| recall | 50% | 47% | 44% | 41% |  |  |  |  |  |  |  |









Conclusion:

* Logistic regression is better as the KNN overfits the data.