Task 3

1) K-Nearest Neighbors (KNN):

```
KNN Classification Report:
                    precision    recall  f1-score   support

    art-et-culture       0.79      0.82      0.80       206
          economie       0.75      0.85      0.80       204
      faits-divers       0.88      0.92      0.90       202
marocains-du-monde       0.81      0.73      0.77       194
            medias       0.87      0.86      0.87       197
           orbites       0.62      0.56      0.59       178
         politique       0.68      0.68      0.68       198
           regions       0.77      0.79      0.78       184
           societe       0.69      0.66      0.68       213
             sport       0.97      0.98      0.97       214
         tamazight       0.92      0.90      0.91       210

          accuracy                           0.80      2200
         macro avg       0.80      0.80      0.79      2200
      weighted avg       0.80      0.80      0.80      2200

KNN Accuracy: 0.7990909090909091
Accuracy for tamazight: 0.8952380952380953
Accuracy for orbites: 0.5617977528089888
Accuracy for faits-divers: 0.9158415841584159
Accuracy for marocains-du-monde: 0.7319587628865979
Accuracy for societe: 0.6619718309859155
Accuracy for sport: 0.9766355140186916
Accuracy for economie: 0.8529411764705882
Accuracy for politique: 0.6818181818181818
Accuracy for medias: 0.8629441624365483
Accuracy for regions: 0.7880434782608695
Accuracy for art-et-culture: 0.8203883495145631
```

Precision: Precision measures the accuracy of positive predictions made by the model. For each class, it is the ratio of true positive predictions to the total number of instances predicted as positive (true positive + false positive). For example, for the class "faits-divers," the precision is 0.88, indicating that 88% of instances predicted as "faits-divers" were actually true positives.

Recall (Sensitivity or True Positive Rate): Recall measures the ability of the model to correctly identify positive instances out of all the actual positive instances in the dataset. For each class, it is the ratio of true positive predictions to the total number of actual positive instances in the dataset (true positive + false negative). For instance, for the class "tamazight," the recall is 0.90, indicating that the model correctly identified 90% of the actual instances of "tamazight."

F1-score: The F1-score is the harmonic mean of precision and recall. It provides a single score that considers both precision and recall, making it useful when there is an uneven distribution between precision and recall. For each class, it is calculated as 2 * (Precision * Recall) / (Precision + Recall).

Support: The support represents the number of instances in each class in the test set. For example, the class "art-et-culture" has a support of 206, meaning there are 206 instances in the test set belonging to this class.

Accuracy: Accuracy is the overall correctness of the model's predictions and is calculated as the ratio of correctly classified instances to the total number of instances in the test set. In this case, the overall accuracy of the model is 0.80, meaning it correctly classified 80% of the instances in the test set.

Macro Avg: The macro average calculates the average of the metrics (precision, recall, F1-score) for each class independently. It gives equal weight to each class, regardless of class imbalance. In this case, the macro average F1-score is 0.79.

Weighted Avg: The weighted average calculates the weighted average of the metrics, taking into account the support (number of instances) for each class. It provides a more balanced view when classes have different sizes. In this case, the weighted average F1-score is also 0.80.

Enhancements: we can change at the number of neighbors and change at hyper parameters

---

2)Using Random Forest Algorithm:

I used ensemble method to increase the accuracy of the model and to reduce the overfitting .

```
Random Forest Classification Report:
                   precision    recall  f1-score   support

    art-et-culture       0.79      0.91      0.85       206
          economie       0.71      0.87      0.78       204
       faits-divers       0.88      0.94      0.91       202
marocains-du-monde       0.80      0.82      0.81       194
            medias       0.94      0.84      0.88       197
           orbites       0.74      0.56      0.64       178
         politique       0.70      0.86      0.77       198
           regions       0.71      0.76      0.73       184
           societe       0.81      0.52      0.63       213
             sport       0.97      1.00      0.99       214
         tamazight       0.94      0.87      0.90       210

          accuracy                           0.82      2200
         macro avg       0.82      0.81      0.81      2200
      weighted avg       0.82      0.82      0.81      2200

Random Forest Accuracy: 0.8154545454545454
Accuracy for tamazight (Random Forest): 0.8666666666666667
Accuracy for orbites (Random Forest): 0.5617977528089888
Accuracy for faits-divers (Random Forest): 0.9356435643564357
Accuracy for marocains-du-monde (Random Forest): 0.8247422680412371
Accuracy for societe (Random Forest): 0.5164319248826291
Accuracy for sport (Random Forest): 1.0
Accuracy for economie (Random Forest): 0.8725490196078431
Accuracy for politique (Random Forest): 0.8585858585858586
Accuracy for medias (Random Forest): 0.8375634517766497
Accuracy for regions (Random Forest): 0.7554347826086957
Accuracy for art-et-culture (Random Forest): 0.9077669902912622
```

Accuracy : 81.5%

Enhancements: we can change at the number of estimators ,change at hyper parameters

---

3) Gradient Boosting Classification:

```
Gradient Boosting Classification Report:
                    precision    recall  f1-score   support

    art-et-culture       0.90      0.85      0.88       206
          economie       0.78      0.84      0.81       204
      faits-divers       0.95      0.89      0.92       202
marocains-du-monde       0.82      0.78      0.80       194
            medias       0.94      0.93      0.94       197
           orbites       0.60      0.65      0.62       178
         politique       0.75      0.73      0.74       198
           regions       0.74      0.80      0.77       184
           societe       0.69      0.64      0.66       213
             sport       0.96      0.98      0.97       214
         tamazight       0.95      0.97      0.96       210

          accuracy                           0.83      2200
         macro avg       0.83      0.82      0.82      2200
      weighted avg       0.83      0.83      0.83      2200

Gradient Boosting Accuracy: 0.8272727272727273
Accuracy for tamazight (Gradient Boosting): 0.9714285714285714
Accuracy for orbites (Gradient Boosting): 0.6460674157303371
Accuracy for faits-divers (Gradient Boosting): 0.8910891089108911
Accuracy for marocains-du-monde (Gradient Boosting): 0.7835051546391752
Accuracy for societe (Gradient Boosting): 0.6384976525821596
Accuracy for sport (Gradient Boosting): 0.9766355140186916
Accuracy for economie (Gradient Boosting): 0.8382352941176471
Accuracy for politique (Gradient Boosting): 0.7323232323232324
Accuracy for medias (Gradient Boosting): 0.934010152284264
Accuracy for regions (Gradient Boosting): 0.8043478260869565
Accuracy for art-et-culture (Gradient Boosting): 0.8543689320388349
```

Accuracy:82%

Enhancements: we can change at the number of estimators ,change at hyper parameters ,Depth of tree

4) SVM Classification:

```
SVM Classification Report:
                    precision    recall  f1-score   support

     art-et-culture       0.90      0.90      0.90       206
           economie       0.83      0.89      0.86       204
       faits-divers       0.97      0.95      0.96       202
marocains-du-monde       0.87      0.87      0.87       194
             medias       0.94      0.91      0.93       197
            orbites       0.66      0.67      0.66       178
          politique       0.80      0.78      0.79       198
            regions       0.76      0.88      0.81       184
            societe       0.77      0.67      0.72       213
              sport       0.99      0.99      0.99       214
          tamazight       0.98      0.94      0.96       210

           accuracy                           0.86      2200
          macro avg       0.86      0.86      0.86      2200
       weighted avg       0.86      0.86      0.86      2200

SVM Accuracy: 0.8613636363636363
Accuracy for tamazight: 0.9428571428571428
Accuracy for orbites: 0.6741573033707865
Accuracy for faits-divers: 0.9455445544554455
Accuracy for marocains-du-monde: 0.865979381443299
Accuracy for societe: 0.6713615023474179
Accuracy for sport: 0.9906542056074766
Accuracy for economie: 0.8872549019607843
Accuracy for politique: 0.7828282828282829
Accuracy for medias: 0.9137055837563451
Accuracy for regions: 0.8804347826086957
Accuracy for art-et-culture: 0.8980582524271845
```

The Best accuracy :86.13%

Enhancements: we can change at the number of estimators ,change at hyper parameters , Experiment with different types of kernels (e.g., linear, polynomial, radial basis function (RBF), sigmoid) to find the one that works best for our data.