# Hand-written Recognition

https://colab.research.google.com/drive/1hVfkZTGiJ8tlqZho68tLj1NFNY-cMpS_

## Overview

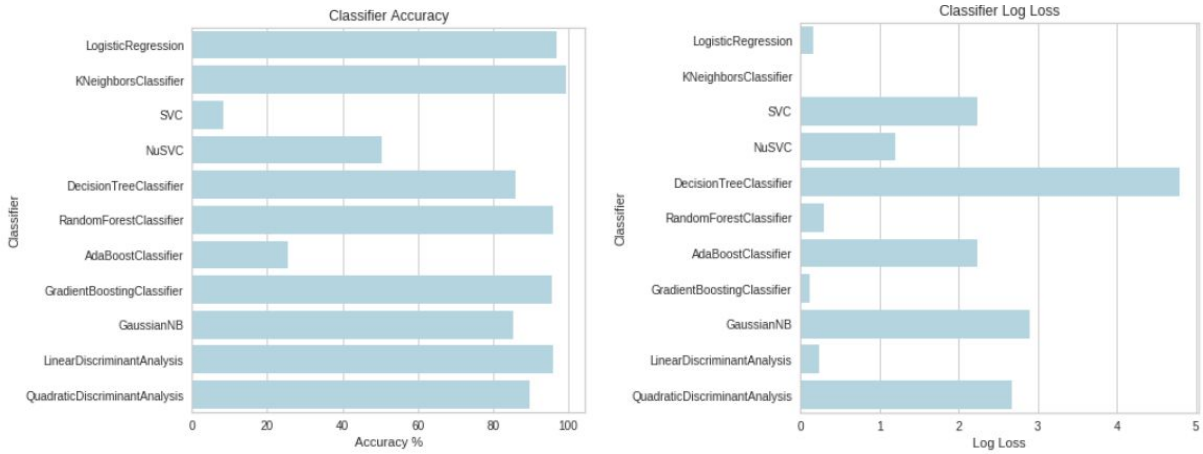
One of the most common uses of machine learning is image recognition. Based on *digits* dataset(which has the same form as a digital transformation from images) in python, this project aims at recognizing hand-written numbers.

## Technique

### *Step1: Model Pre-selection*

Which classifier to choose is one of the most important questions to ask when approaching a machine learning problem. Before going too far with one model, I would like to know which models are good and which are not so that I can save my time with only exploring the good ones. Therefore, I first calculate the model accuracy for 11 of the most common algorithms, which include logistic regression, K-neighbors, etc.

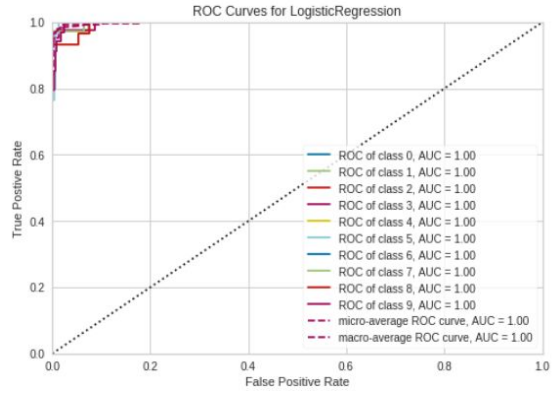
The results are shown below and we can see **KNeighborsClassifier** and **LogisticRegression** are the top-two best models since they have the highest accuracy(98.14% and 96.94% respectively). Accordingly, I explore more details of these two models and compare them from multi-aspects.

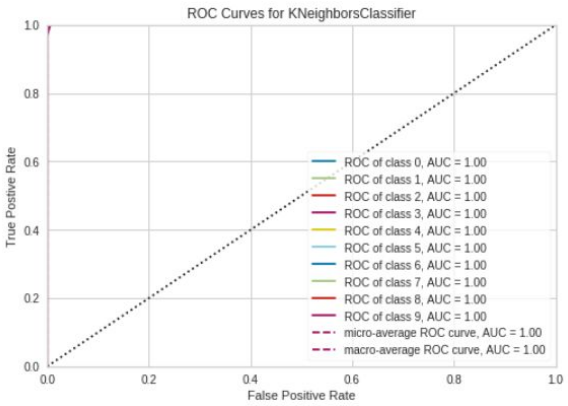


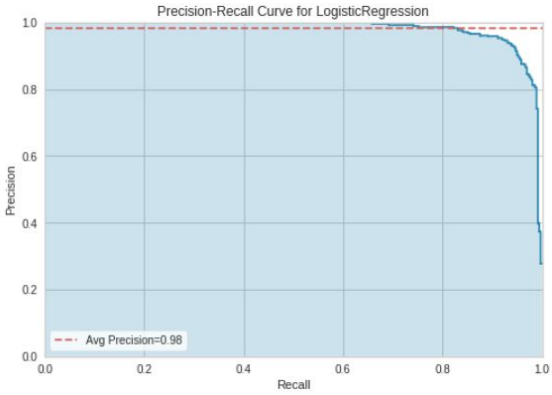

### *Step2: Model Comparison*

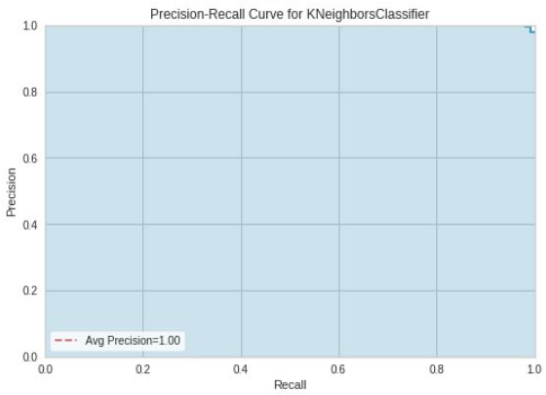
In this project, I compare these two models using three methods: ROC Curve, Precision-Recall Curve and Confusion Matrix. The former two evaluate models in a whole while the Confusion Matrix gives more details about which classes are most easily confused.

From these three corresponding visualizations, we can see that the KNeighborsClassifier outperform the LogisticRegression Classifier as a whole because KNeighborsClassifier has ROC Curves closer to the left top corner, a bigger area under the Precision-Recall curves and fewer values unmatched in the Confusion Matrix.

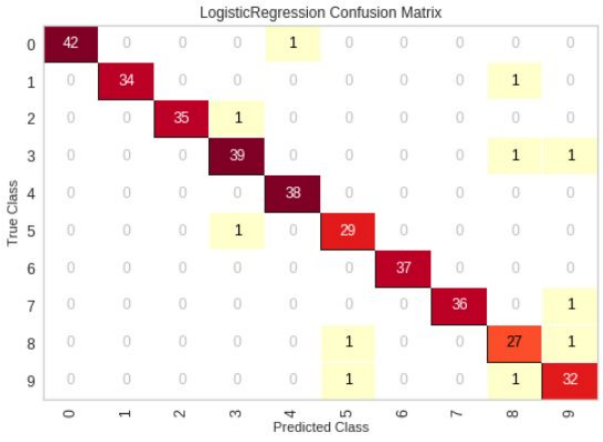


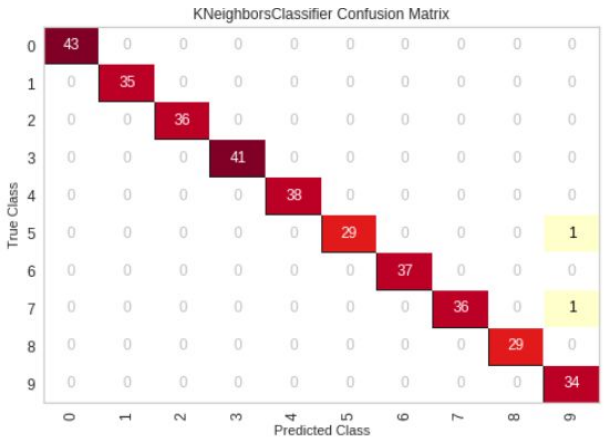




From the Confusion Matrix, we can see the hand-written number “3”, “8” and “9” are the most confusing ones for Logistic Regression since there are more than one unmatches. The hand-written number “5” and “7” are the most confusing ones for K-Neighbors Classifiers.




## Conclusion

With the comparison of two models, we can draw the conclusion that KNeignbors Classifier is the best model to recognize the hand-written numbers given the current dataset. While the model predicts almost all the numbers right, the hand-written number “5” and “7” are still a little confusing for it.