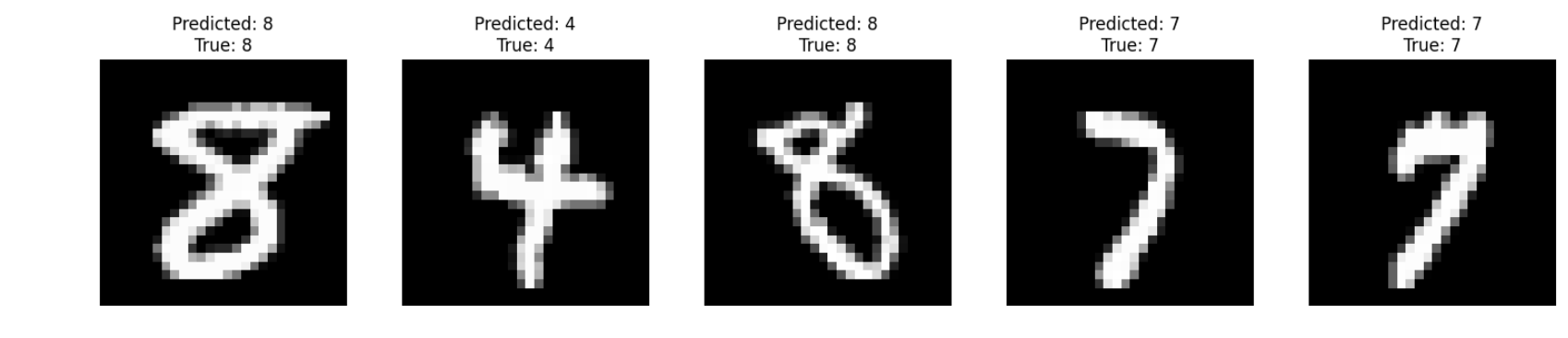
Rayyaan Haamid

Patricia McManus

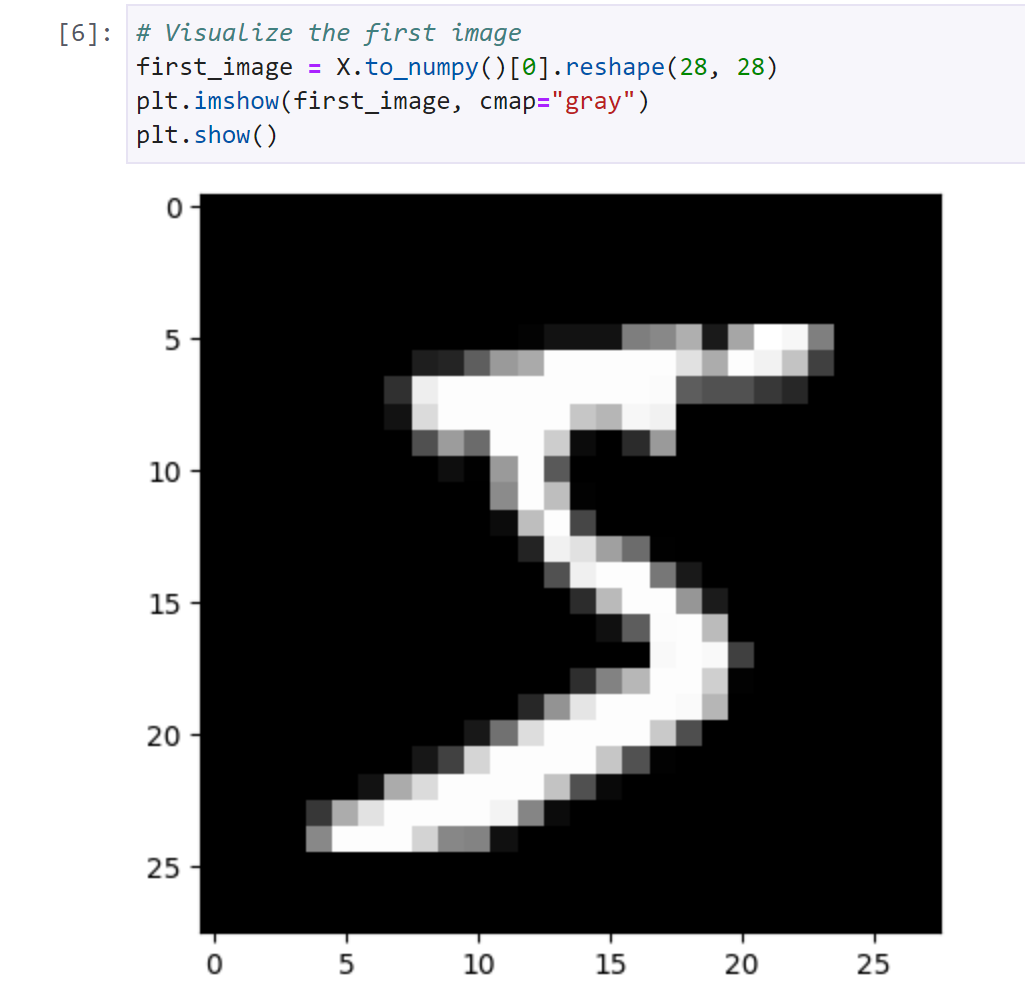
ITAI 1378

January 12 2024

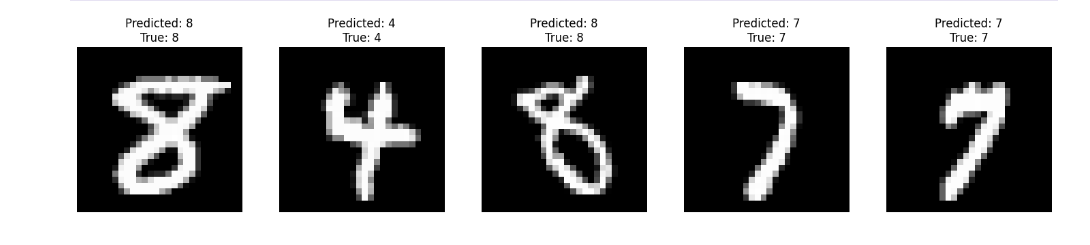
L04 Project Documentation

The Mnist algorithm was to bring up the necessary imaging to allow the K-NN algorithm to predict and show the results of the apparent images by identifying the images and classifying them. In simpler terms, this machine learning algorithm is used for classifying and regression tasks. But for this case, it is not specialized and incredible enough due to that nature is an instant algorithm. And for identifying the images, it uses k nearest neighbors which are essentially the labels for the algorithm in this case. And the change of the k nearest neighbors affects the performance.

Installing and importing the model is always an important thing and will never work without them. Loading the data will give the algorithm what to do with the basis which in this case are the numbers from Mnist. Converting the labels from strings to integers makes sense as models typically need values more than strings as they are not the best values to allow the computer to understand. There is also an example for the Mnist image:



And made it so it splits the data into training and testing sets and then goes into training the model. As this is simple machine learning, it is easily trained and typically has the training data and coding along with making it. with a cute code to allow predictions for the images and after the training, the evaluation for the training comes to light and presents the accuracy of it being 97.13%. this makes sense as it is difficult for AI to perform at 100% as there are human anomalies that are too difficult for AI at the moment to go through but could be changed in the future with neural AI. lastly, the code that makes the algorithm a reality. It has Mnist list of images connected with the k-nn algorithm that takes the images from Mnist and goes through the

Process of what the number is and gives a prediction that is already incredibly high but sometimes could give something wrong and show the true value for each.

As how simple of a model it is, having issues with it is minimal and luckily didn’t have many concerns with it. and the performance of the model is pretty high with the knowledge of the accuracy of the model. For something harder and vague typically has issues with accuracy but this had none of them, while it is possible for the algorithm to be better with the accuracy but is good enough for something to notice the 3% and correct that themselves and not do that for the latter.

This program is not bad for the reason of its creation, having a high accuracy, being able to take in all the data from Mnist and process that quickly. This could be because of the studio lab but it has a good speed of processing the whole model with ease and being able to be simple for beginners to understand. But the cons of the model are not being able to understand letters, and not being able to process other datasets with potentially better data as it needs many changes to adapt to other datasets. The simplicity of the model is quite apparent with its simple use of machine learning but it gives an understanding of what it is and is good at what it is doing. To make the algorithm go faster than its raw version would remove all the unnecessary labels that come with the datasets so using techniques like “feature selection to retain only the most informative features. Changing the data to lower-dimensional space to reduce the variety of data as K-NN is not the best with high dimensions. Approximating the nearest neighbors could be a good way to increase the speed but sacrificing the accuracy of the result. The use of hyperparameter k is crucial as having too low will allow flexibility but is prone to overfitting and a larger k will allow nicer boundaries. Getting all the necessary features and removing unnecessary or noisy features.

Reference list:

Brownlee, J. (2019, August 6). How to Avoid Overfitting in Deep Learning Neural Networks. MachineLearningMastery.com

Tokuç, A. A. (2022, November 9). k-Nearest Neighbors and High Dimensional Data. Baeldung on Computer Science

Nithyashree, V. (2021, July 19). Step-by-Step guide for Image Classification on Custom Datasets. Analytics Vidhya3