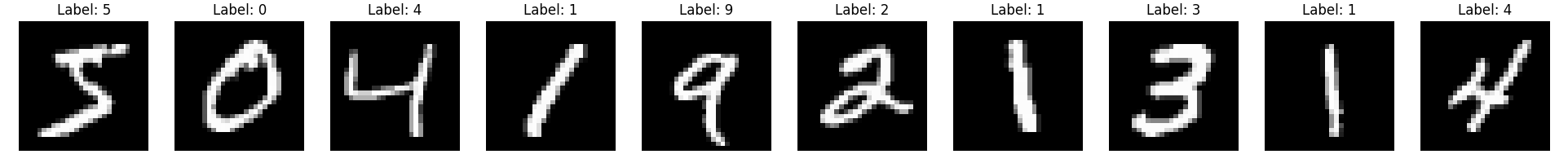
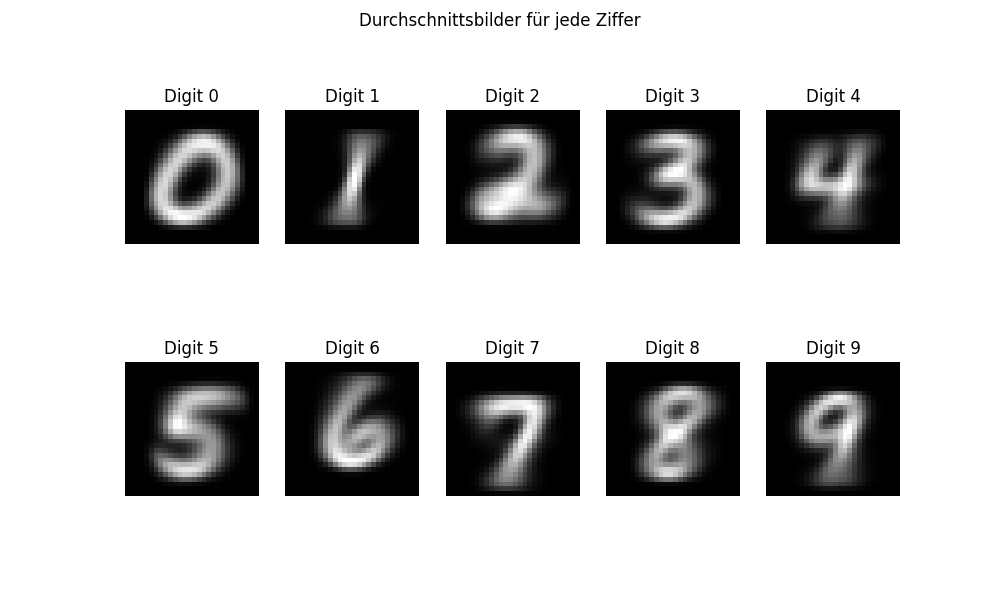
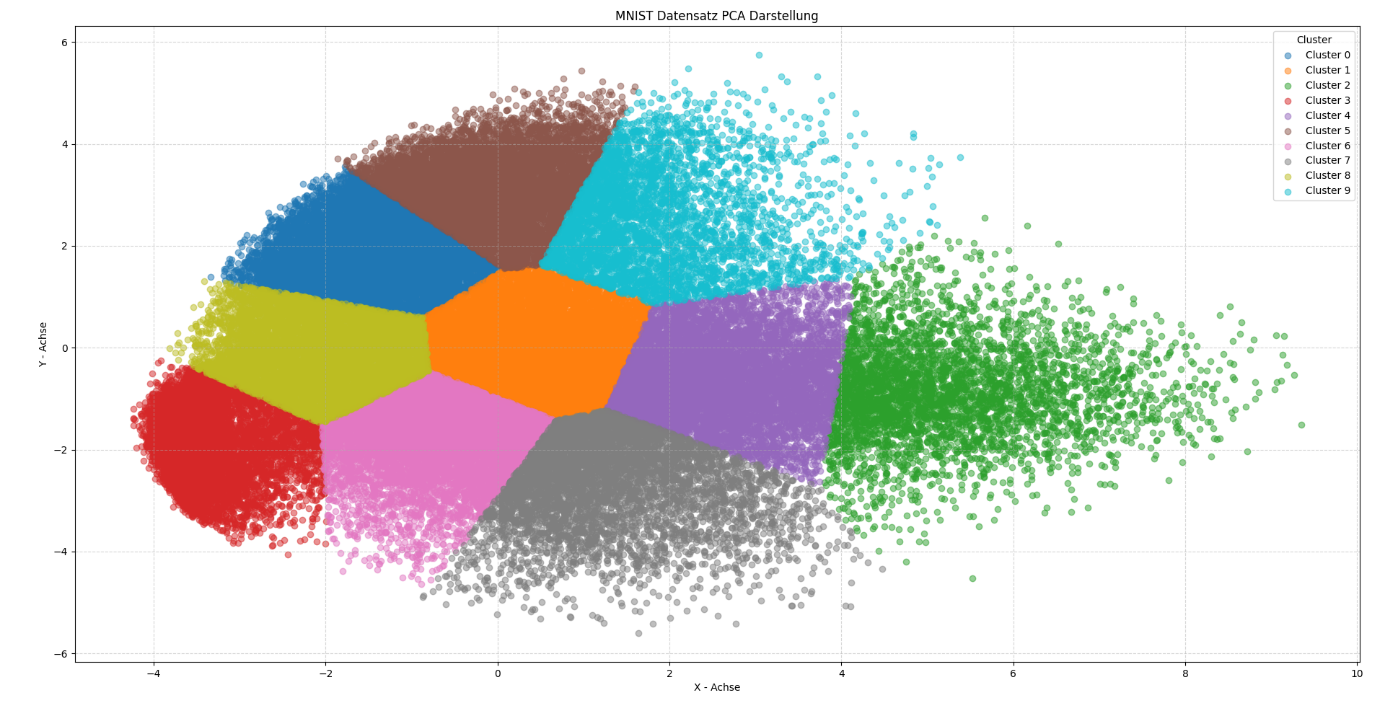
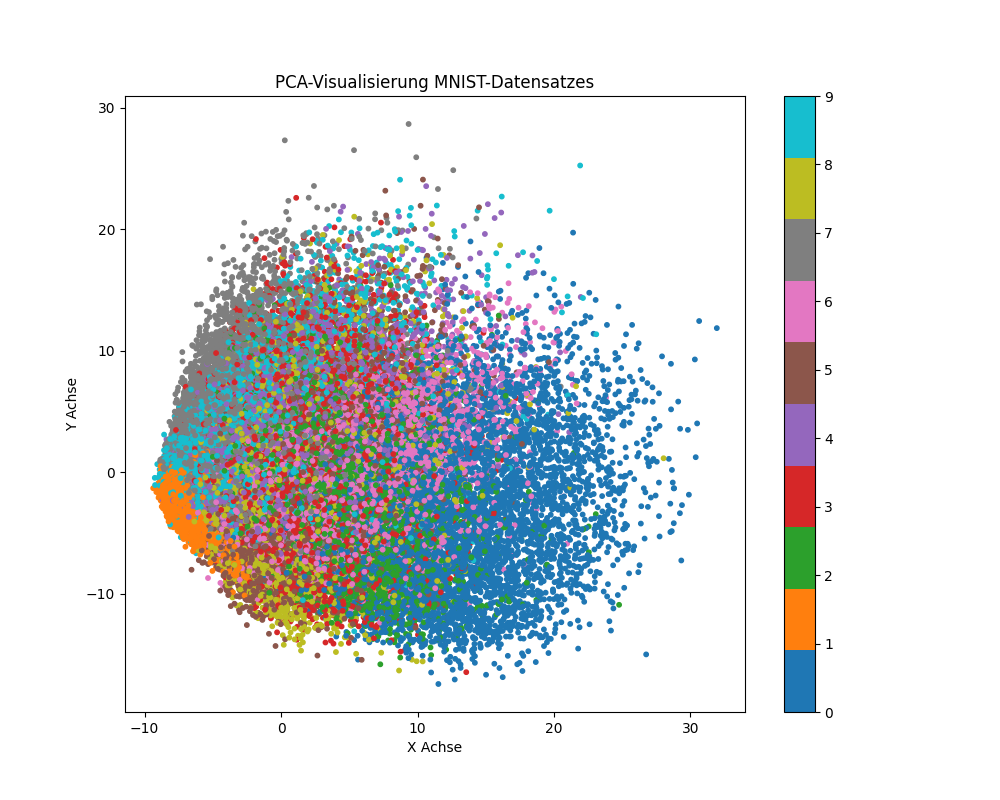
Der MNIST (Modified National Institute of Standards and Technology database) Datensatz ist ein zahlenbasierter Datensatz mit knapp 70.000 Daten. Inhalt der daten sind handschriftlich geschriebene Zahlen von 0 bis 9 in 28x28 Pixel-Bildern, die auf individuelle Art und Weise aufgeschrieben worden sind.  

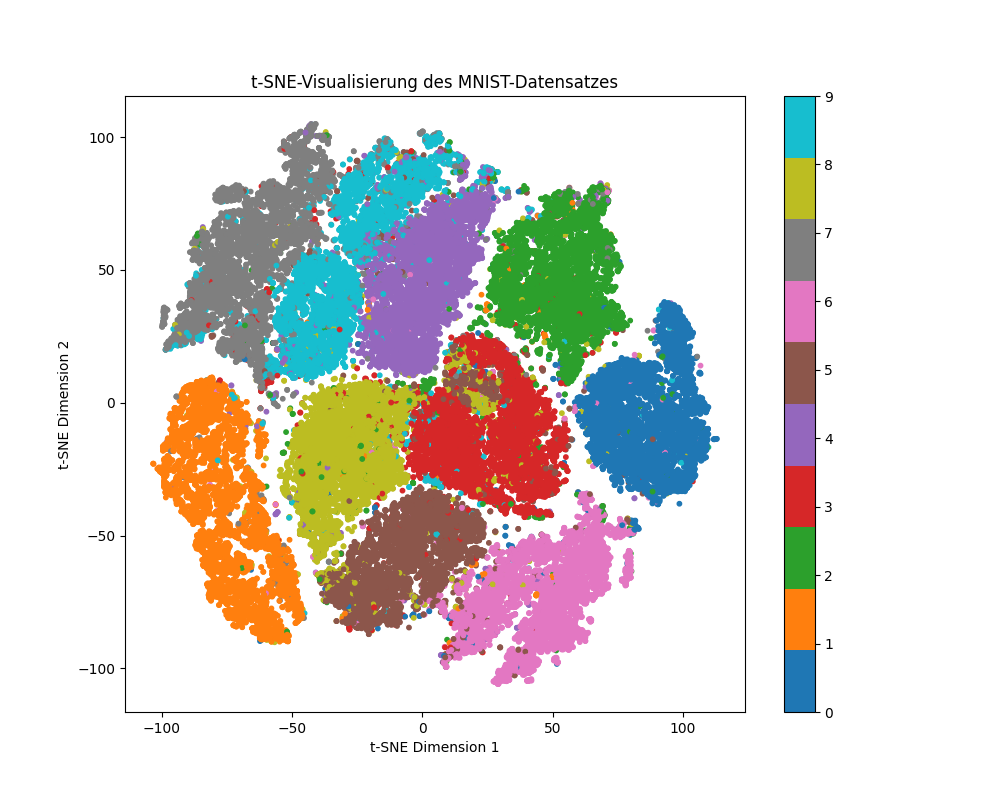
Von diesen 70.000 Daten sind 60.000 Trainingsdaten und 10.000 Testdaten. Die 60.000 Trainingsdaten sind, wie der Name schon verrät, da um das Modell zu trainieren. Die 10.000 Bilder in den Testdaten sind dafür da, um die trainierten Daten zu vergleichen. Ziel ist es, mit Hilfe von neuronalen Netzen die Zahlen richtig zu erkennen und zu klassifizieren. Mithilfe von diesem freizugänglichen Datensatz können Neuronale Netze erstellt werden, dadurch das maschinelle Lernen im Allgemeinen gefördert werden. Das Training und die Evaluierung von Algorithmen für die optische Zeichenerkennung (OCR) und das maschinelle Lernen verwendet.

Auch kann man diese Daten in verschiedensten Varianten visuell aufzeigen. In den folgenden Bildern wurden die Daten in beispielsweise Cluster, PCA und t-SNE Visualisierung, da diese unserer Meinung nach die besten sind. Mit dem Cluster sollen die Daten gruppiert werden und unterschiedlich zu anderen Gruppen (Clustern) gemacht werden. 

Desweiteren wurde eine Prinicpal Compotent Analysis (PCA) Darstellung erzeugt. Die PCA soll Date mit hohen bzw. vielen Dimensionen sollen damit verstädnlicher gemacht werden indem man die Dimemsionen reduziert. Dabei sollen viele Varioatienen beibehalten werden. PCA transformiert die Daten in einen neuen Merkmalsraum, der durch die Hauptkomponenten definiert wird. Diese Hauptkomponenten sind die linearen Kombinationen der ursprünglichen Merkmale, die die maximale Varianz im Datensatz erklären.



Und als letztes wurde der Datensatz mithilfe der t-Distributed Stochastic Neighbor Embedding (t-SNE) Methode visualisiert. Auch diese Methode wird genutzt, um die hochdimensionalen Räume zu vereinfachen bzw. diese zu reduzieren. Aber im Gegensatz zum PCA Prinzip, wird hier versucht die Merkmale im gleichen Datenraum beizubehalten.



<https://docs.ultralytics.com/de/datasets/classify/mnist/>

<https://www.tensorflow.org/datasets/catalog/mnist>