**Problem Statement:**

To design a project from the given MNIST dataset to digitally classify based on the given label using the SVM algorithm.

**Packages And IDE’s:**

**Google Colaboratory:**

It is a jupyter notebook environment that runs in the browser using google cloud servers. It is a widely popular cloud service for machine learning that features free access to faster GPU and TPU computing resources. Colab notebooks allow you to combine executable code and rich text in a single document, along with images, HTML, laTeX, and more. It is simply a cloud-based service that replicates jupyter notebook in the cloud. Colaboratory can be used to perform many tasks such as writing and run code, create its associated documentation, and display graphics. It allows for faster GPU access and easy sharing. It is highly integrated with google drive making them easy to set up and access.

**Sklearn:**

Scikit-learn popularly known as Sklearn is a free software machine learning library for Python programming. It is a simple and efficient tool for predictive data analysis. It features various classification, regression and clustering algorithms such as random forests, gradient boosting, support vector machines, k-means, and DBSCAN and it is designed to interoperate with the Python numerical and scientific library NumPy and scipy. It is open-source and is commercially usable to everyone.

**Pandas:**

Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. Pandas allow for various data manipulation operations such as merging, reshaping, selecting as well as data cleaning and data wrangling features. It is a free software that is mainly used for data analysis.

**Problem Solution:**

The given problem is a type of classification problem which comes under supervised learning. The given problem is solved by using the support vector classifier(SVC) algorithm, which is a type of support vector machine(SVM) algorithm that is used in solving classification problems.

The objective of a Linear SVC (Support Vector Classifier) is to fit the data you provide, returning a "best fit" hyperplane that divides or categorizes our data.

The SVM algorithm is implemented in practice using a kernel. A kernel transforms an input data space into the required form. SVM uses a technique called the kernel trick. Here, the kernel takes a low-dimensional input space and transforms it into a higher-dimensional space. In other words, you can say that it converts nonseparable problems into separable problems by adding more dimension to it. It is most useful in non-linear separation problems. The kernel trick helps you to build a more accurate classifier. The kernel type used in our model is “linear”.

Linear Kernel A linear kernel can be used as a normal dot product for any two given observations. The product between two vectors is the sum of the multiplication of each pair of input values.

The pandas(a python data analysis library) library is used to load the given dataset from the given CSV file using the read\_csv method.

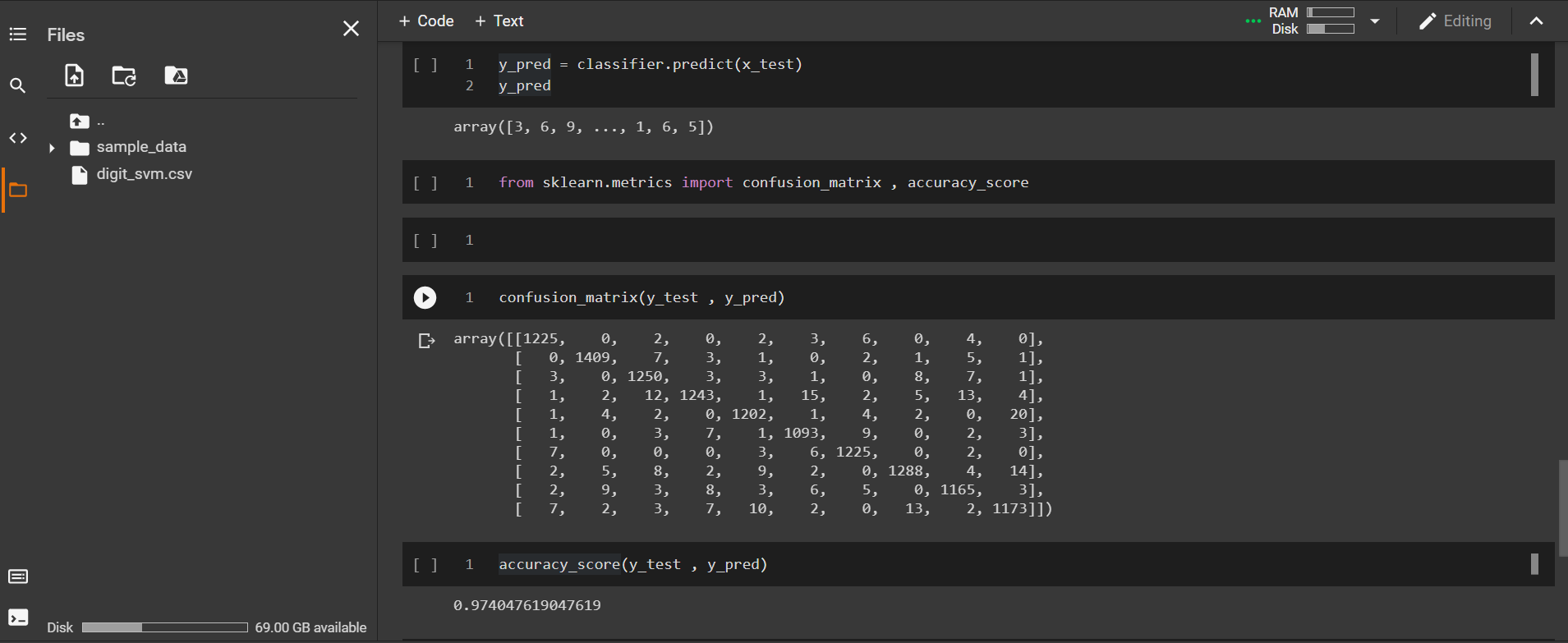
The Scikit-Learn(a machine-learning library) library is used for building the prediction model. The given columns are divided into two types of variables – dependent and independent variables. Then the given dataset is divided into a training set and a test set in the ratio of 7:3 using the test\_train method from the model\_selection class. The logistic regression classifier object is created using the LogisticRegression class. Build your model on the training set using the fit method and then perform prediction on the test set using the predict method.

The accuracy score of the prediction model is found by using the accuracy\_score method on a scale of 0 to 1. This method notifies how accurate our prediction model is.

The confusion matrix of the prediction model is found by using the confusion\_matrix method. A confusion\_matrix is a table that is used to evaluate and visualize the performance of a classification model. The fundamental of a confusion matrix is that the number of correct and incorrect predictions are summed up class-wise.

Both accuracy\_score and confusion\_matrix are methods in the metrics class in the scikit-learn library.

**Code:**

****

**Conclusion:**

The ML model has an accuracy of nearly 98% and the ML model can be effectively used to predict the number label using the given pixel values for the greyscale image label. The above-mentioned ML model is successfully implemented in python in google colaboratory as a notebook file.