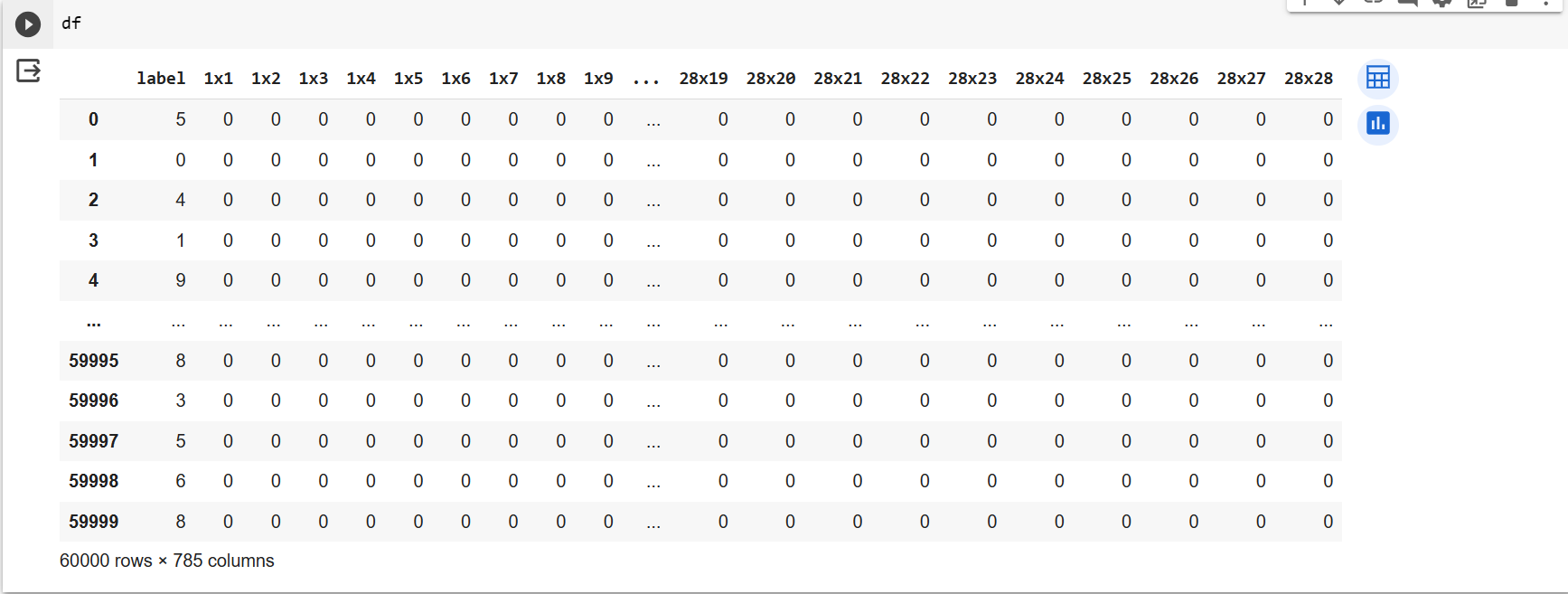
Image Classification ML Project

# Data Set:

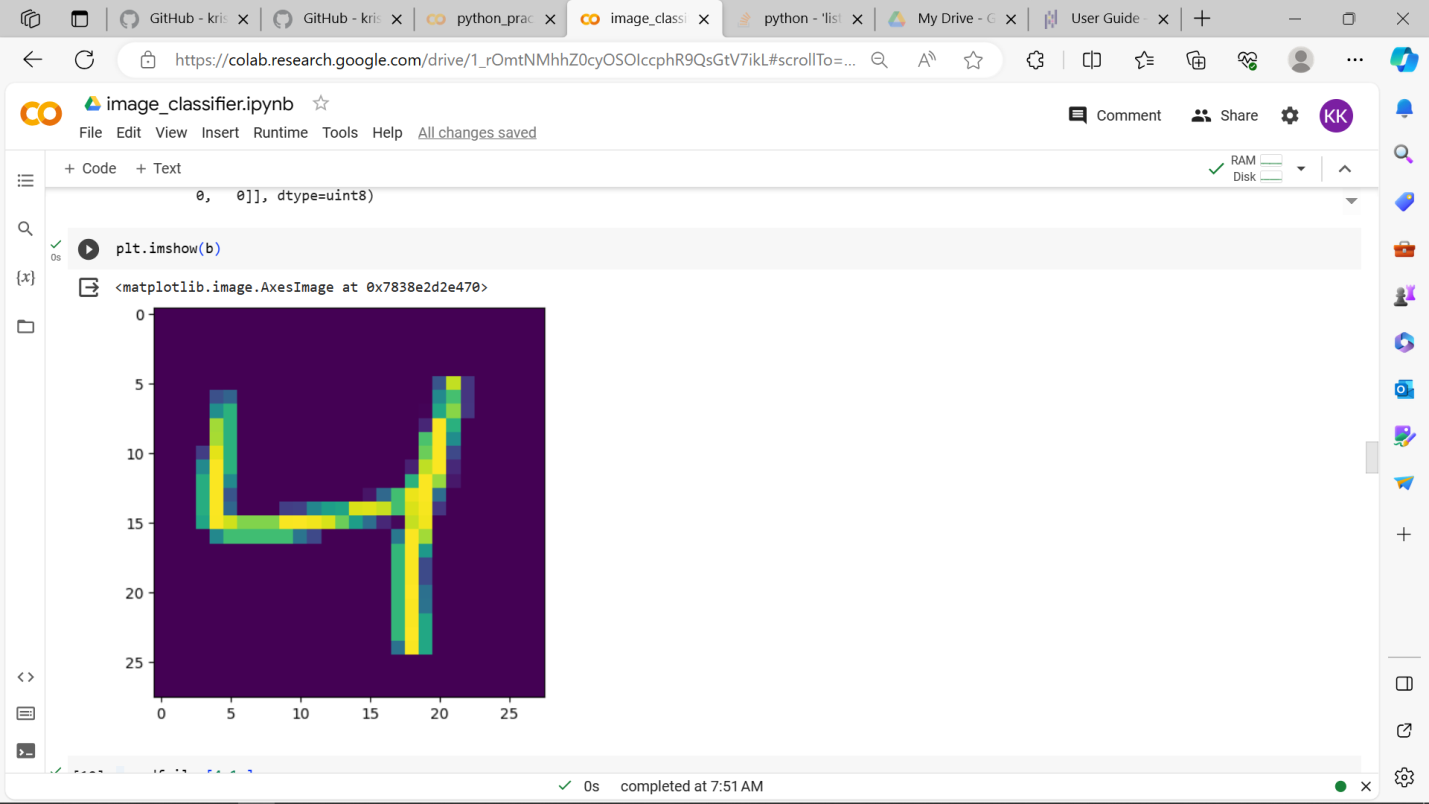
Data Set was downloaded from an external data source site (kaggle)

Name: mnist.csv

About: it’s a set of images that contain handwritten numbers in pixel format (28x28), and has 60000 rows × 785 data



Plotting a sample row from the above data:



ML Model:

Using **RF(Random Forest Classifier)**

A Random Forest Classifier is a popular machine learning algorithm that falls under the ensemble learning category. It is used primarily for classification tasks, which involve categorizing data into predefined classes or categories. The algorithm is known for its robustness, flexibility, and ability to deliver high-quality results in a wide range of applications.

Here are the key characteristics and components of a Random Forest Classifier:

**1. Ensemble Learning**:

* Random Forest is an ensemble learning method, meaning it combines the predictions from multiple machine learning models to improve overall performance. In this case, it combines multiple decision trees.

**2. Decision Trees**:

* The fundamental building blocks of a Random Forest are decision trees. A decision tree is a tree-like structure where each internal node represents a feature, each branch represents a decision or rule based on that feature, and each leaf node represents a class label.

**3. Randomness**:

* Randomness is a core element in Random Forest. It introduces variability in the model-building process, making each tree slightly different. This randomness can help reduce overfitting and increase the model's robustness.

**4. Bootstrap Aggregating (Bagging)**:

* Random Forest employs a technique called bagging, which involves creating multiple subsets (samples) of the training data through random sampling with replacement. Each decision tree is trained on one of these subsets.

**5. Feature Randomness**:

* In addition to sampling data, Random Forest introduces randomness in feature selection. When building each tree, it considers only a random subset of the available features. This is done to increase diversity among the trees and reduce the risk of overfitting.

**Advantages of Random Forest Classifier**:

1. **High Accuracy**: Random Forest is known for its high accuracy in classification tasks. It is particularly robust when dealing with noisy data and complex relationships.
2. **Reduced Overfitting**: The combination of bagging and feature randomness helps reduce overfitting, making Random Forest a reliable choice even for small datasets.
3. **Handles Mixed Data Types**: Random Forest can handle both numerical and categorical data, making it versatile for various applications.
4. **Scalability**: It can efficiently handle large datasets and a high number of features.
5. **Outlier Robustness**: Random Forest is less sensitive to outliers and anomalies compared to some other algorithms.

Overall, Random Forest Classifiers are a valuable tool in the machine learning toolkit, offering a balance between accuracy and robustness for a wide range of classification tasks

