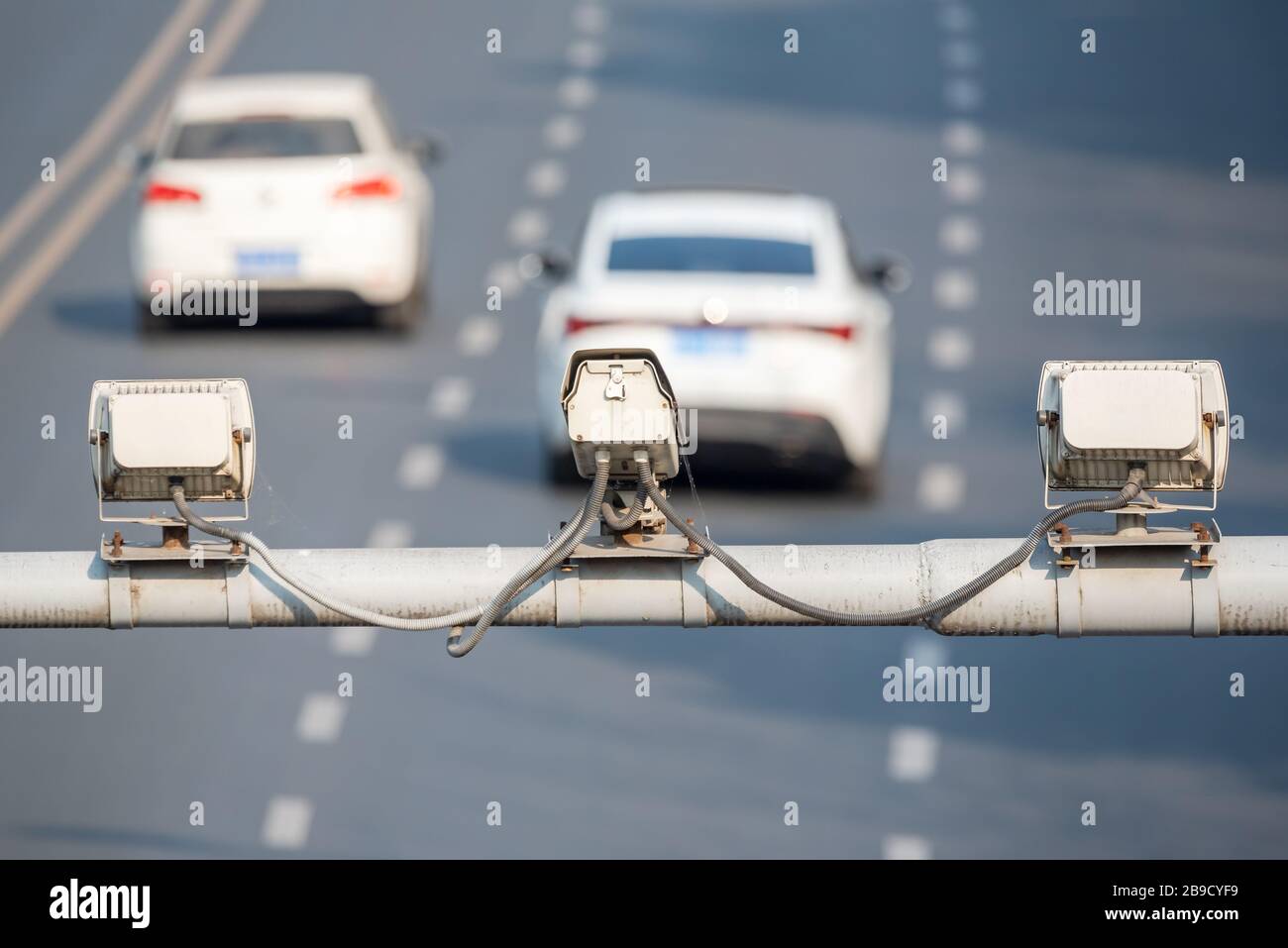
Text, letter

Description automatically generated

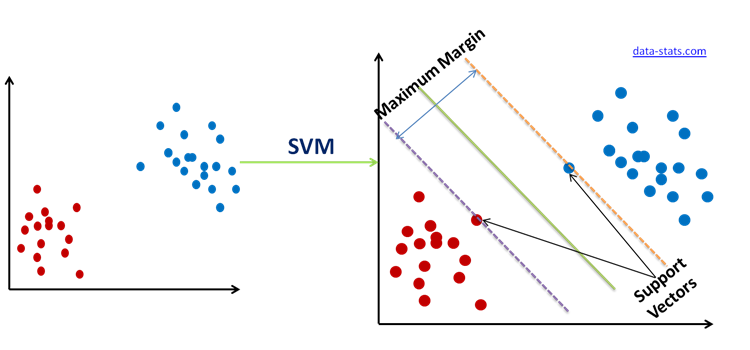


Ever wondered how speed cameras detect the plates of  
a vehicle accurately?

This operation is done with machine learning technique  
called digit recognition. To simplify how digit recognition  
works, you create an AI model and feed it with data and label your data with useful and meaningful names and then insert new data without telling the device it’s label, at this moment the model should predict what is the label of the data you inserted.

The dataset we are currently using is imported from the [MNIST](https://en.wikipedia.org/wiki/MNIST_database) Database and we have taken 1797 images from it.  
We’ve split the data into 2 parts , 80% training and 20% testing because these percentages provide the best results for our experiment.

The algorithm used to predict numbers in our project is called SVM.

***Support Vector Machine*** are supervised models with associated learning algorithms that analyze data for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, SVM maps training examples to points in space to maximize the width of the gap between the two categories. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

Chart, scatter chart

Description automatically generatedHyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. If the number of input features is 2, then the hyperplane is just a line. If the number of input features is 3, then the hyperplane becomes a two-dimensional plane. It becomes difficult to imagine when the number of features exceeds 3.

To separate the two classes of data points, there are many possible hyperplanes that could be chosen. Our objective is to find a plane that has the maximum margin, i.e., the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.Chart, scatter chart

Description automatically generated

**Coding implementation:**

1. First of all we load our digits from the MNIST Database our dataset is full of numbers that vary between 0 -> 9 so we started off by calculating the mean and found out that it’s value is equal to 4.49081803 which is exactly between 0 and 9. After that we were curious to figure out which number is repeated the most “ Mode “ and the graph below shows that number 3 is the mode of this set with the score of 183.

Bar Chart :

Chart, bar chart

Description automatically generated

Histogram :

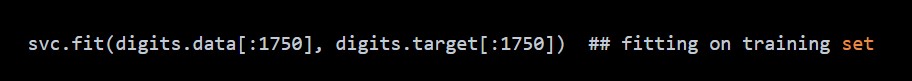
Chart, histogram

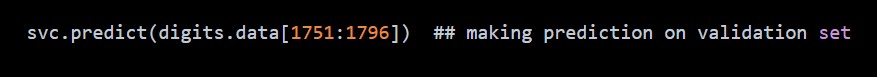
Description automatically generated

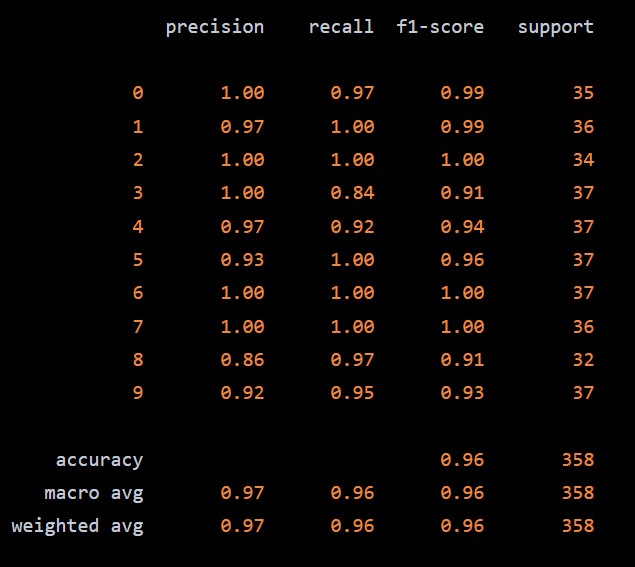
The next part shows how the machine learning concepts works with our idea, these are few steps that we follwed to train and test our model :

1. We create a model to feed it with our algorithm   
   Graphical user interface

   Description automatically generated with medium confidence  
   This code activates the SVM algorithm into our model .
2. Feed the model with the training data

  
  
This fills our model with the current known data, at this step the model studies the pattern of every number entered.

1. Predict the result of the untrained data  
     
   This functions uses the previous known data and analyses it to predict the unknown data .
2. Last but not least show a heatmap A picture containing diagram

   Description automatically generated  
   This heatmap shows the error percentage visually
3. Finally there is a classification report that shows accuraccy and other data numerically  
   As shown here the accuracy percentage is 96% which is a high precentage .

In the end, digit recognition is a really useful technique and is used widely across the globe in different aspects .

As a conclusion to our project , there are some points that we figured out that are considered main keys :

1. Graphical user interface, application, Teams

   Description automatically generated Our Ai model is very sensetive to the gamma values.  
     
   The gamma parameter defines how far the influence of a single training example reaches .  
    with low values meaning “far” and high values meaning “close”.  
     
     
   - If the gamma is low (far) it takes in consideration the further points which makes it more smooth and linearly .

* If the gamma is high ( close ) it takes in consideration the closest points to the hyperplane make it more crooked .
* the C parameter is a regularization parameter used to set the tolerance of the model to allow the misclassification of data points in order to achieve lower generalization error.

Another point related to this topic as well is the type of function used with this algorithm .  
There are 3 types of functions used with our algorithm which are “RBF” , “Linear” , “Polynomial”   
We tested each of them and these are the results :

* RBF:

Table

Description automatically generated with low confidence

* Poly:



* Linear:



2- The second point is about the svm margins , as shown in previous images about svm’s we explained how margins work and after a lot of experimenting we found out that the more we maximize our margins the higher the accuracy we achieve in the future .

Sources used in this project :

* [Hackernoon.com](https://hackernoon.com/building-handwritten-digits-recognizer-using-support-vector-machine-fos3wqs)
* [WikiPedia](https://en.wikipedia.org/wiki/MNIST_database)