

Hand Written Digit Recognition

PROJECT REPORT

SUBMITTED BY

Harsh Gupta

ROLL NO- 2014670

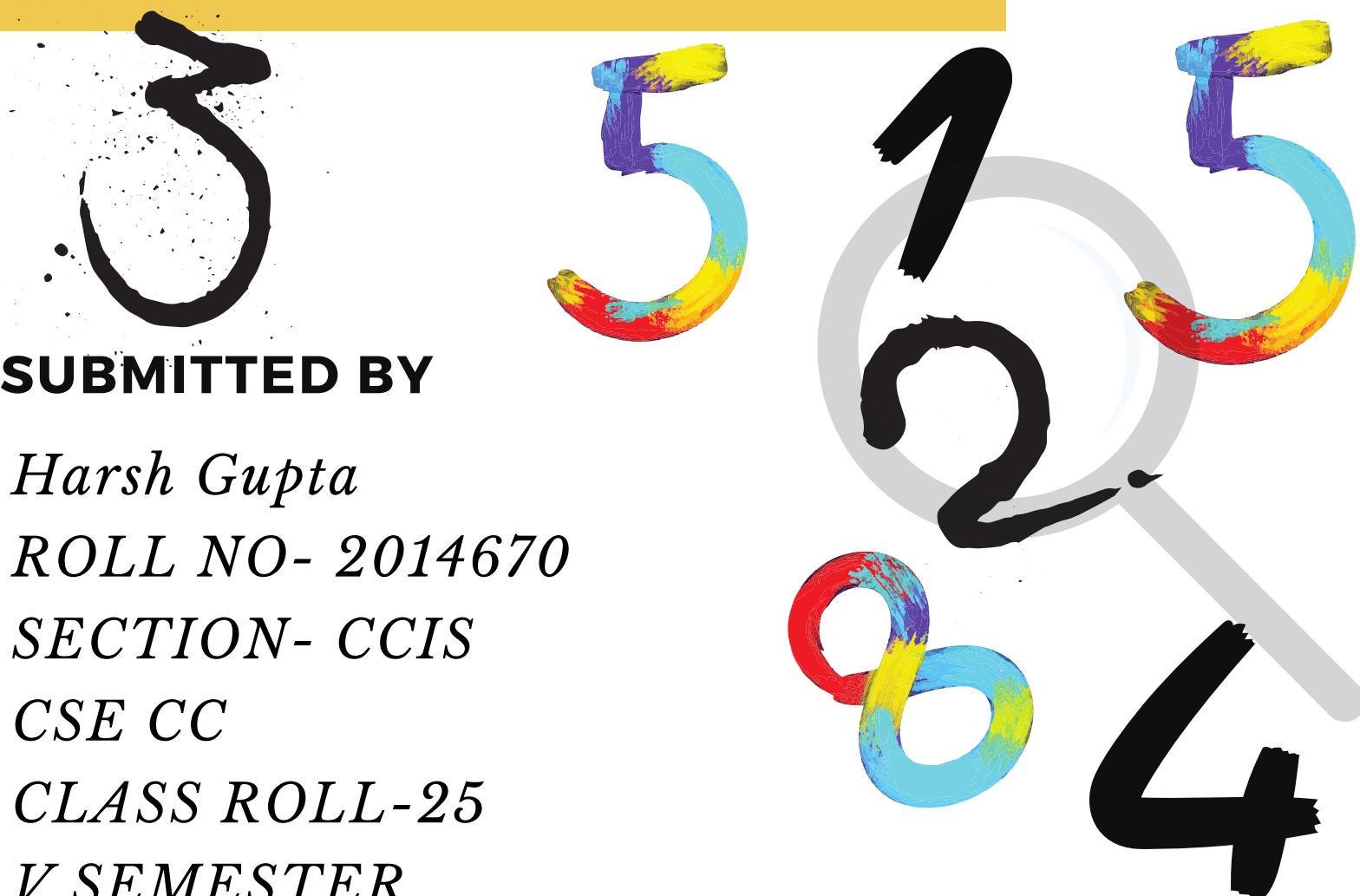
SECTION- CCIS

CSE CC

CLASS ROLL-25

V SEMESTER

Email id- Harshc2303@gmail.com



PROBLEM STATEMENT



Input



Prediction

take a handwritten input and
successfully predict which digit is
that

ABOUT PROJECT

Introduction

Hand-written digit recognition is the capacity of a PC to perceive the human written by hand digits from various sources like pictures, papers, contact screens, and so forth, and order them into 10 predefined classes (0-9).

Working on Hand Written dataset is just like the hello world for testing different machine learning algorithms.

With the adaptation of machine learning, there has been a generous measure of innovative work that has given a flood to profound learning and AI alongside man-made reasoning. With time, machines are getting increasingly modern, from working out the fundamental calculations to doing retina recognition have made our lives more safer and manageable.

In Handwritten digit recognition, we face many difficulties in view of various styles of composing of various people groups as it is certifiably not an Optical character recognition.

Digit recognition has numerous applications like number plate acknowledgment, postal mail arranging, bank check processing, and so on.

Libraries used

Different scikit-learn libraries like NumPy, matplotlib, pandas, Sklearn and seaborn have been utilized for the execution reason. First and foremost, we will download the MNIST datasets, trailed by stacking it and perusing those CSV documents utilizing pandas.

- `matplotlib.pyplot`

`matplotlib.pyplot` is a collection of functions that make `matplotlib` work like MATLAB. Each `pyplot` function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc.

csv

CSV (Comma Separated Values) is a simple file format used to store tabular data, such as a spreadsheet or database.

- `pandas`

Pandas is an open-source Python package that is most widely used for data science/data analysis and machine learning tasks. It is built on top of another package named Numpy, which provides support for multi-dimensional arrays.

- `sklearn`

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistent interface in Python.

cv2

OpenCV is a great tool for image processing and performing computer vision tasks. It is an open-source library that can be used to perform tasks like face detection, objection tracking, landmark detection, and much more.

DATASET USED

MNIST DATASET

Handwritten character recognition is an expansive research area that already contains detailed ways of implementation which include major learning datasets, popular algorithms, features scaling and feature extraction methods.

MNIST dataset (Modified National Institute of Standards and Technology database) is the subset of the NIST dataset which is a combination of two of NIST's databases: Special Database 1 and Special Database 3. Special Database 1 and Special Database 3 consist of digits written by high school students and employees of the United States Census Bureau, respectively. MNIST contains a total of 70,000 handwritten digit images (60,000 - training set and 10,000 - test set) in 28x28 pixel bounding box and anti-aliased. All these images have corresponding Y values which apprises what the digit is.

MNIST dataset (Modified National Institute of Standards and Technology database) is the subset of the NIST dataset which is a mix of two of NIST's information bases: Special Data set 1 and Special Database 3. Exceptional Database 1 and Exceptional Database 3 comprise digits composed by secondary school understudies and workers of the United States Census Bureau, individually. MNIST contains a sum of 70,000 written by hand digit pictures (60,000 - preparing set and 10,000 - test set) in 28x28 pixel jumping box and hostile to associated. This multitude of pictures has comparing Y esteems which informs what the digit is.

STEPS FOR COMPLETE EXECUTION

step 1

import all the necessary libraries first to use them like pandas matplotlib sklearn etc.

step 2

load the MNIST data set either from the local storage or from the sklearn library.

step 3

load the MNIST data set either from the local storage or from the openml datasets.

```
In [38]: from sklearn.datasets import fetch_openml  
from sklearn.model_selection import train_test_split  
from sklearn.linear_model import LogisticRegression  
from sklearn import metrics  
import numpy as np  
import matplotlib.pyplot as plt  
import pandas as pd  
%matplotlib inline
```

```
In [39]: mnist= fetch_openml('mnist_784')
```

step 4

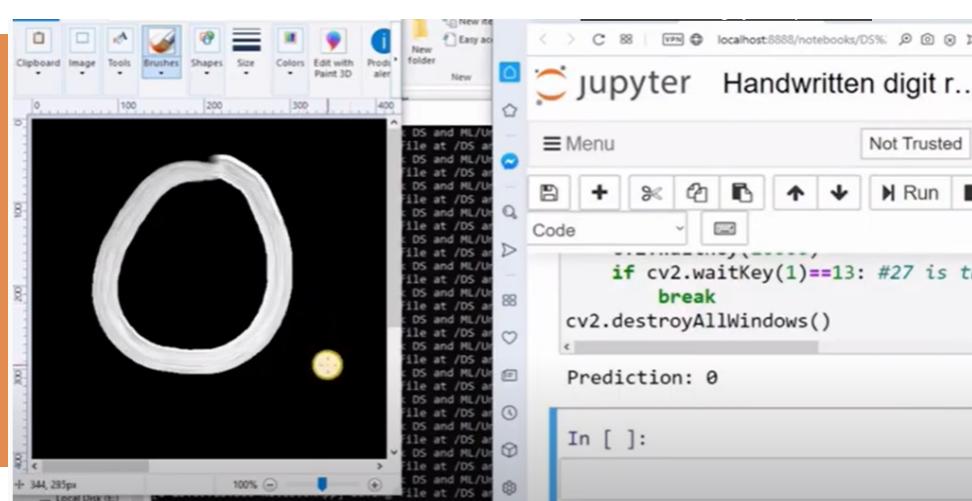
Now we will read the dataset in our variables to further split the data into Training_data, Testing_data for our machine learning algorithm to work on .

step 5

Now we will calculate the accuracy of the model with the help of testing_data against the model with the training_data. and print the accuracy.

step 6

Now as we have the prediction accuracy we can test our input digits with the same accuracy and generate the output of the predicted machine.



TRAINING AND TESTING DATA?

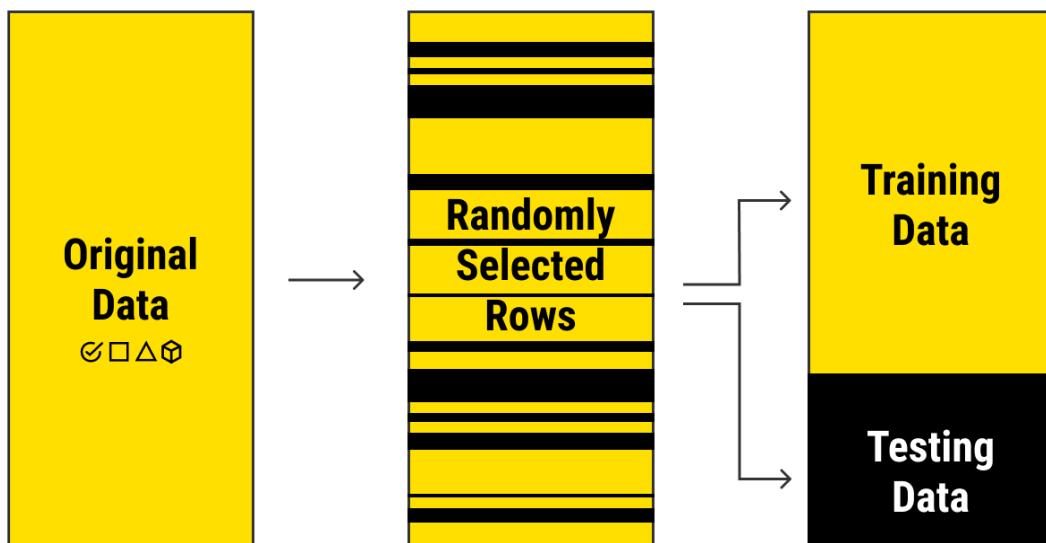
WHAT IS TRAINING DATA?

TRAINING DATA IS A SET OF LARGE DATA USED TO TEACH THE MACHINE LEARNING MODEL FOR SUPERVISED MACHINE LEARNING THE TRAINING

DATA IS LABELLED AND FOR UNSUPERVISED LEARNING THE TRAINING DATA IS UNLABELLED. IT HELPS THE MACHINE TO PRODUCE SOPHISTICATED RESULTS.

80%
TRAINING DATA

20%
TESTING DATA



ACCURACY IS THE RATE AT WHICH THE TEST DATA IS CORRECTLY PREDICTED

WHAT IS TESTING DATA?

TESTING DATA IS A COMPARATIVELY SMALLER SET OF DATA USED FOR TESTING THE MACHINE LEARNING MODEL IT HELPS TO CALCULATE THE ACCURACY OF THE MODEL. TRAINING DATA IS GENERALLY SPLIT INTO 20 % AND 80% WHERE THE SMALLER ONE IS KEPT FOR TESTING AND THE BIGGER DATA SET IS USED FOR TRAINING.

CONCLUSION

In this project, we have implemented a machine learning model for handwritten digit recognition using MNIST datasets, based on deep and machine learning algorithms.

Future enhancements

Future enhancements of these machine learning projects are boundless as we get more and more advanced algorithms and more public application access to Machine learning models there are endless enhancements

References

“Handwriting recognition”:

<https://en.wikipedia.org/wiki/Handwritingrecognition>

[2] “What can a digit recognizer be used for?”:

<https://www.quora.com/What-can-a-digit-recognizer-be-used-for>

[3] ”Handwritten Digit Recognition using Machine Learning Algorithms”,
S M Shamim, Mohammad Badrul Alam Miah, Angona Sarker, Masud
Rana & Abdullah Al Jobair.

[4] "Hand Written Digit Recognition" By IG tech team

[5] OpenML to fetch and search the right Dataset for my Model

[6] Handwritten Digit Recognition using Machine and Deep Learning
Algorithms Report by-

- Ritik Dixit Computer Science and Engineering Acropolis Institute of Technology & Research Indore, India dixitritik17@gmail.com
-
- Rishika Kushwah Computer Science and Engineering Acropolis Institute of Technology & Research Indore, India rishikakushwah99@gmail.com
-
- Samay Pashine Computer Science and Engineering Acropolis Institute of Technology & Research Indore, India samaypashine7@gmail.com
-