LUCOSA

LAIKIPIA UNIVERSITY

Computer Students Association

Developer Student Clubs

INTRO TO MACHINE LEARNING (ML)

BY MARTIN OYWA (CHAIRPERSON LUCOSA)

Twitter: @martinoywa
Github: @martinoywa
Instagram: @lucosa_ke

OVERVIEW

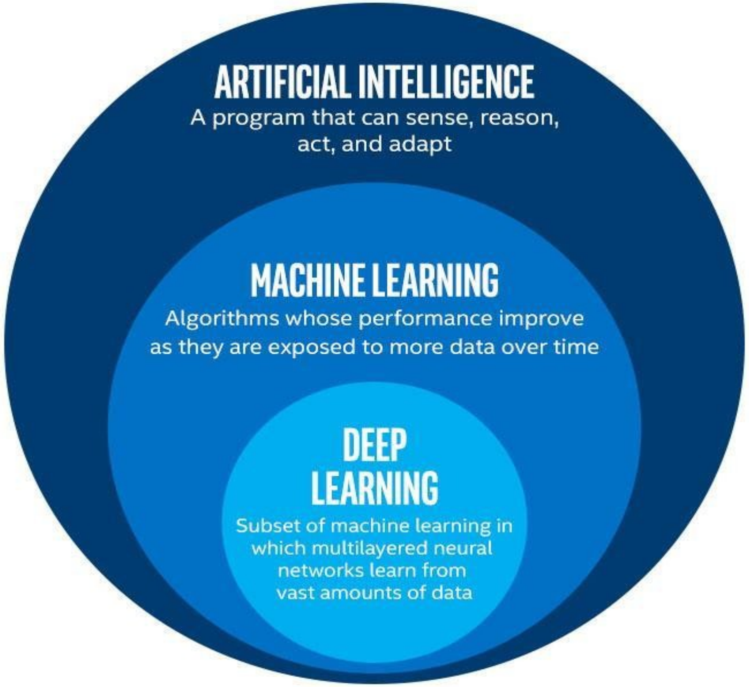
- WHAT IS ML?
- CATEGORIES OF ML
- APPLICATIONS OF ML
- TOOLS FOR ML
- WHAT ARE DATASETS?
- ENVIRONMENT SETUP
- BUILD A SIMPLE MODEL (ADMIT/NOT ADMIT) (PANDAS, NUMPY, SKLEARN)
- BUILD A MORE COMPLEX MODEL (DIGIT CLASSIFIER) (TENSORFLOW (TF))

WHAT IS MACHINE LEARNING (ML)?

- ARTIFICIAL INTELLIGENCE – A BRANCH OF COMPUTER SCIENCE DEALING WITH THE SIMULATION OF INTELLIGENCE USING COMPUTERS.

- MACHINE LEARNING – THE STUDY AND CONSTRUCTION OF PROGRAMS THAT ARE NOT EXPLICITLY PROGRAMMED, BUT LEARN PATTERNS AS THEY ARE EXPOSED TO MORE DATA OVER TIME.

- DEEP LEARNING – MACHINE LEARNING THAT INVOLVES USING VERY COMPLICATED MODELS CALLED “DEEP NEURAL NETS”

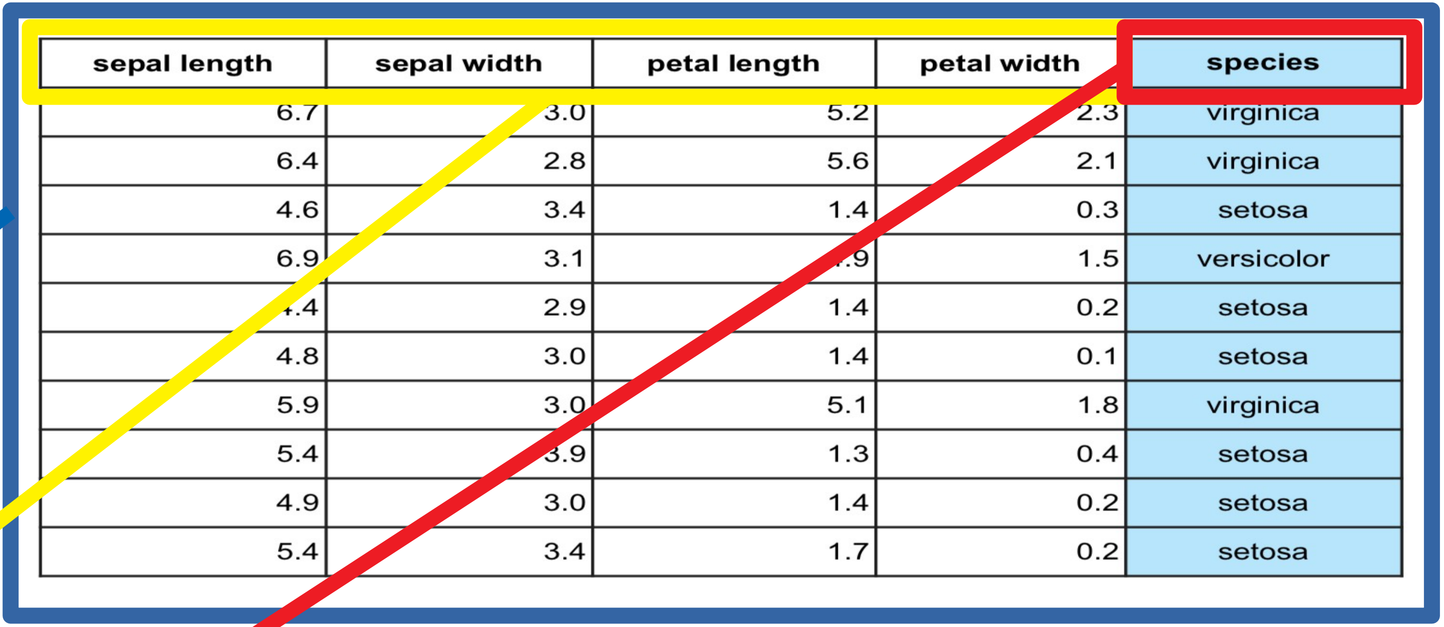


ARTIFICIAL INTELLIGENCE
A program that can sense, reason, act, and adapt

MACHINE LEARNING
Algorithms whose performance improve as they are exposed to more data over time

DEEP LEARNING
Subset of machine learning in which multilayered neural networks learn from vast amounts of data

MACHINE LEARNING TERMINOLOGY



sepal length	sepal width	petal length	petal width	species
6.7	3.0	5.2	2.3	virginica
6.4	2.8	5.6	2.1	virginica
4.6	3.4	1.4	0.3	setosa
6.9	3.1	4.9	1.5	versicolor
4.4	2.9	1.4	0.2	setosa
4.8	3.0	1.4	0.1	setosa
5.9	3.0	5.1	1.8	virginica
5.4	3.9	1.3	0.4	setosa
4.9	3.0	1.4	0.2	setosa
5.4	3.4	1.7	0.2	setosa

- DATASET: DATA COMMONLY IN ROWS AND COLUMNS

- FEATURES: ATTRIBUTES OF DATA

- TARGET: COLUMN TO BE PREDICTED

CATEGORIES/ TYPES OF ML

THERE ARE TWO MAIN TYPES OF MACHINE LEARNING:

- SUPERVISED LEARNING: HERE THE MODEL IS TRAINED TO MAKE PREDICTIONS USING DATA WITH A TARGET COLUMN.
- UNSUPERVISED LEARNING: THE MODEL DOES NOT HAVE A TARGET COLUMN, THUS IT HAS TO FIND STRUCTURE IN THE DATA BY ITSELF.

APPLCATIONS OF ML

FRAUD DETECTION

(IDENTIFY FRAUDULENT CREDIT CARD TRANSACTIONS.)

FEATURE EXAMPLES:

- TRANSACTION TIME
- TRANSACTION AMOUNT
- TRANSACTION LOCATION
- CATEGORY OF PURCHASE

THE ALGORITHM COULD LEARN WHAT FEATURE COMBINATIONS SUGGESTS UNUSUAL ACTIVITY.

OTHER APPLCATIONS OF ML

HEALTH

INDUSTRIAL

FINANCE

TRANSPORT

TOOLS FOR ML

FRAMEWORKS:

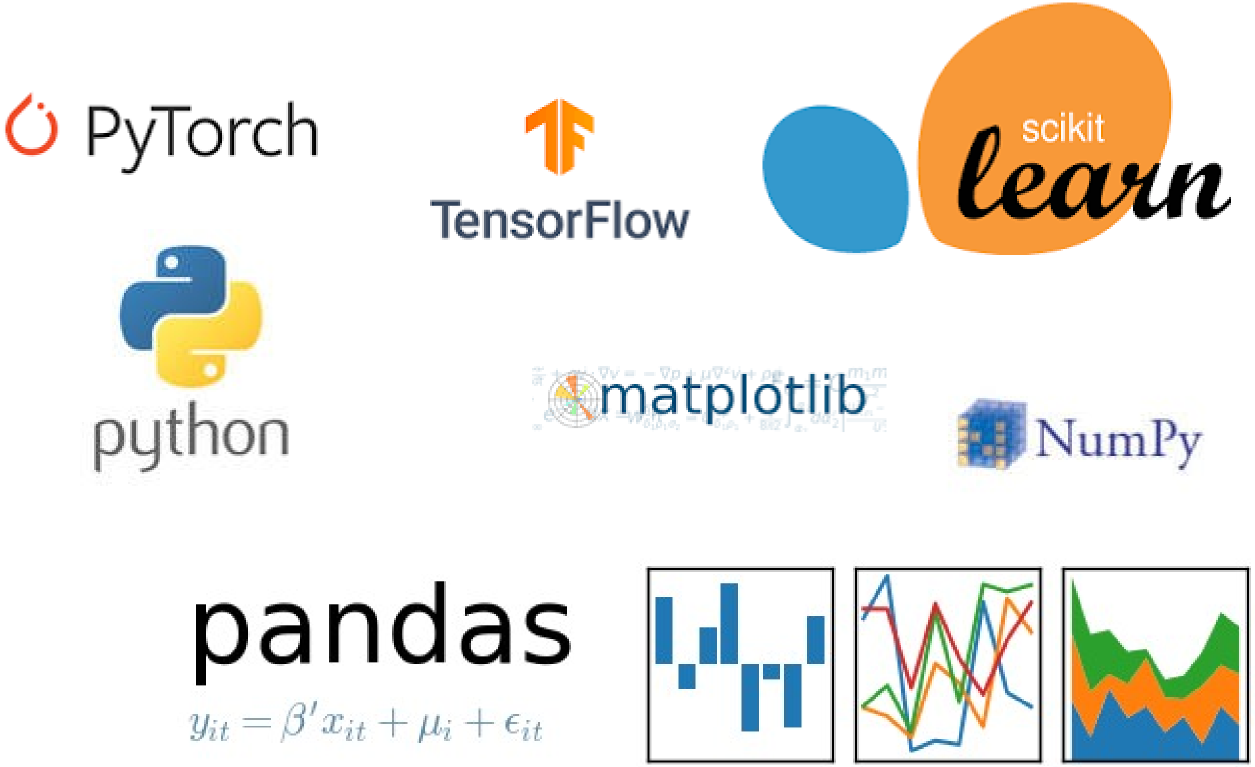
- PYTORCH
- TENSORFLOW
- SCIKIT LEARN

PROGRAMMING LANGUAGES:

- PYTHON

OTHER TOOLS:

- NUMPY
- PANDAS
- MATPLOTLIB



WHAT ARE DATASETS

A DATASET IS JUST A FANCY NAME FOR DATA THAT IS NEEDED FOR AI.

COMMON FORMS OF DATASETS:

- TEXT (FOR NATURAL LANGUAGE PROCESSING (NLP))
- CSV (FOR ALMOST ANYTHING)
- IMAGES (FOR COMPUTER VISION)

ENVIRONMENT SETUP

WHAT IS AN ENVIRONMENT? AN ENVIRONMENT IN SIMPLE TERMS IS SPACE WHERE YOU GET TO DO YOUR DEVELOPMENT FROM.

THE MOST COMMON TYPE OF ENVIRONMENT IS THE VIRTUAL ENVIRONMENT. WHICH IS AN ABSTRACTION WHEREBY YOUR CODE AND ITS DEPENDENCIES CANNOT INTERFERE WITH OTHER ENVIRONMENTS.

TOOLS THAT ARE USED TO CREATE ENVIRONMENTS INCLUDE, VIRTUALENV, PIPENV, AND CONDA (ANACONDA).

ENVIRONMENT SETUP CONT.

DOWNLOAD ANACONDA https://www.anaconda.com/distribution/#download-section

WHY ANACONDA? ANACONDA COMES WITH MOST OF THE TOOLS / PACKAGES NEEDED FOR ML.

HOW TO INSTALL ANACONDA:

- DOWNLOAD LATEST PYTHON INTERPRETER HERE (For Windows) https://www.python.org/downloads/
- WINDOWS…. RUN THE anaconda.exe FILE AND FOLLOW THE STEPS.
- FOR LINUX / UBUNTU … OPEN TERMINAL (ctrl+alt+T), TYPE cd /downloads, TYPE chmod +x anaconda.sh, FINALLY ./anaconda.sh

ENVIRONMENT SETUP CONT.

INSTALLING OTHER PACKAGES

WINDOWS:

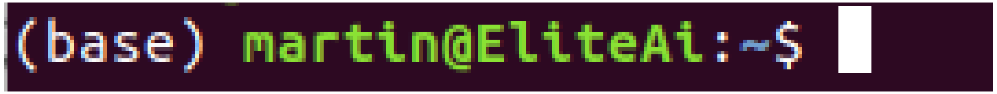
- OPEN ANACONDA PROMPT CMD (CLICK START AND TYPE anaconda promt cmd)
- YOU'LL GET A CMD WITH (base).

ENVIRONMENT SETUP CONT.

INSTALLING OTHER PACKAGES

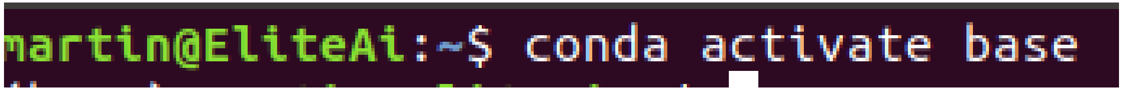
UBUNTU / LINUX:

- OPEN TERMINAL (ctrl+alt+T)
- You’re terminal should have the (base) like this



```
(base) martin@EliteAi:~$ 
```

- If not, type this on the prompt

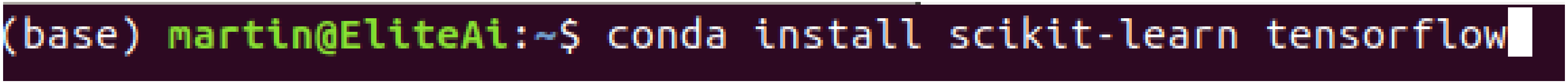


```
martin@EliteAi:~$ conda activate base
```

ENVIRONMENT SETUP CONT.

INSTALLING OTHER PACKAGES

- conda install scikit-learn tensorflow



```
(base) martin@EliteAi:~$ conda install scikit-learn tensorflow
```

DEMO TIME

ADMIT / NOT ADMIT

WE’LL BE BUILDING A SIMPLE MODEL TO DETERMINE WHETHER WE SHOULD ADMIT A STUDENT OR NOT.

BACKGROUND

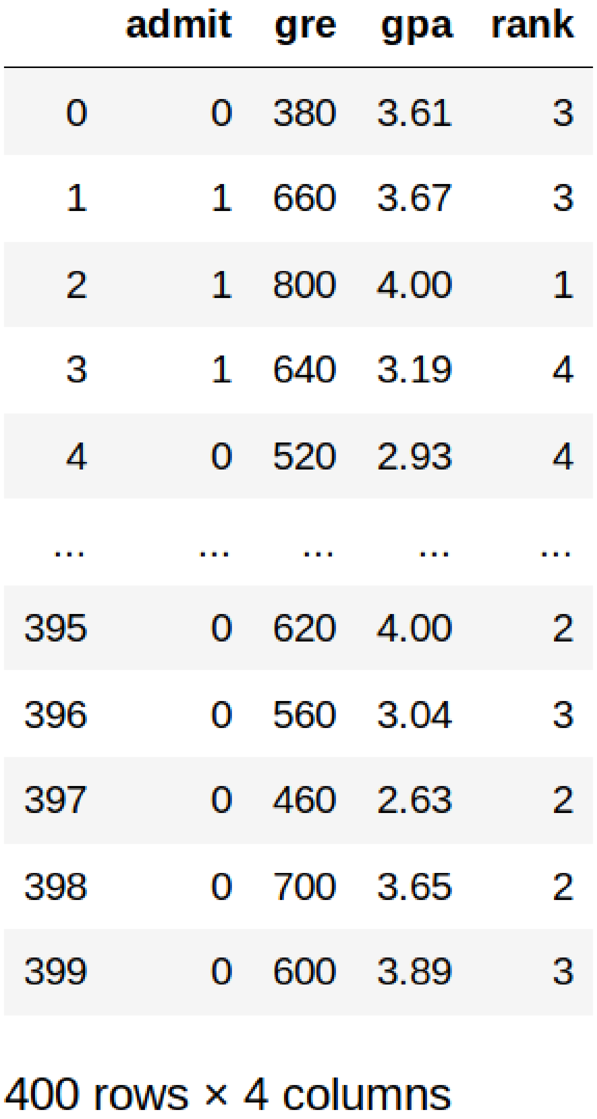
THE ADMISSION PROCESS IS USUALLY A REPETITIVE AND CUMBERSOME TASK THAT REQUIRES A LOT OF TIME AND EFFORT FROM PEOPLE. SO HOW CAN WE MAKE THE PROCESS EASIER? IS THERE A WAY WE CAN USE MACHINE LEARNING TO TRAIN A MODEL TO DO THE WORK FOR US?

ADMIT / NOT ADMIT

SAMPLE DATASET

FEATURES:
- GRE
- GPA
- RANK

TARGET:
- ADMIT



	admit	gre	gpa	rank
0	0	380	3.61	3
1	1	660	3.67	3
2	1	800	4.00	1
3	1	640	3.19	4
4	0	520	2.93	4
...	...	...	...	...
395	0	620	4.00	2
396	0	560	3.04	3
397	0	460	2.63	2
398	0	700	3.65	2
399	0	600	3.89	3

400 rows × 4 columns

ADMIT / NOT ADMIT

WHAT TOOLS WE’LL USE:

- PANDAS
- NUMPY
- SCKITI-LEARN

CODE (INTRO TO ML/AdmitNotAdmit/):

- INTRO TO ML.html
- INTRO TO ML.ipynb (NOTEBOOK)

DIGIT CLASSIFIER

FOR THIS CASE, WE’LL BE BUILDING A MORE COMPLEX MODEL TO CLASSIFY DIGITS FROM 0-9 (USING A DATASET CALLED MNIST).

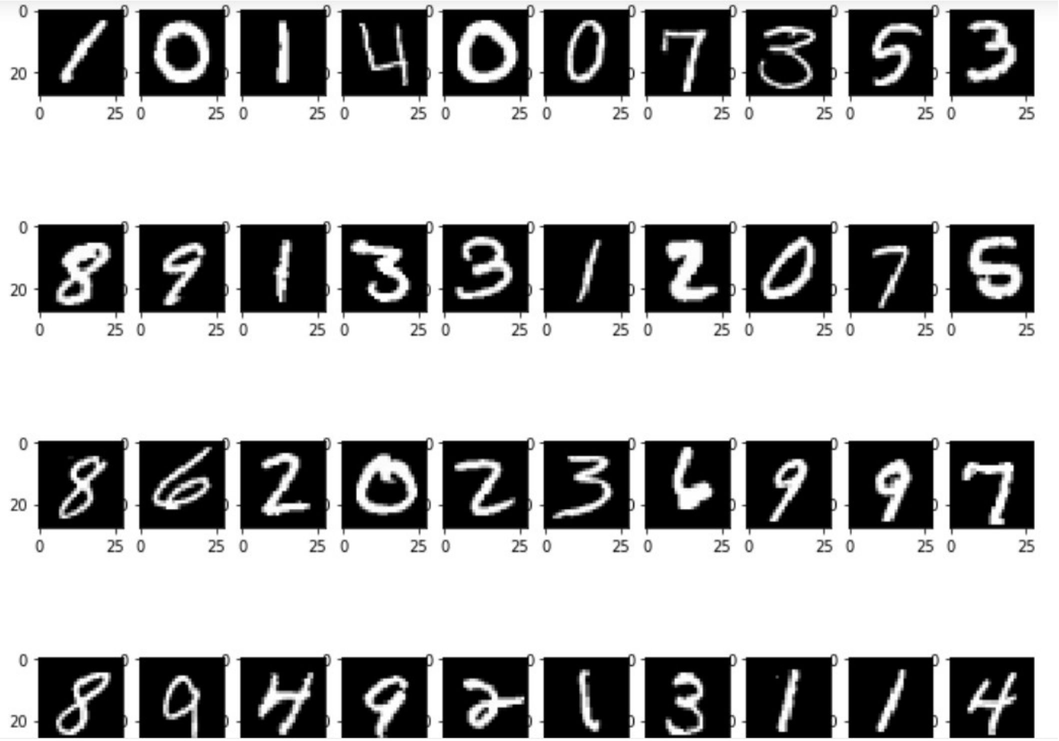
USE CASES

- RECOGNIZING ADDRESSES ON STREETS
- STUDY TOOLS (TEACH YOUR KID NUMBERS)
- INTRODUCTION TO COMPUTER VISION

DIGIT CLASSIFIER

SAMPLE MNIST DATASET (THE IMAGES)

IMAGES ARE USUALLY 28x28 IN SIZE.

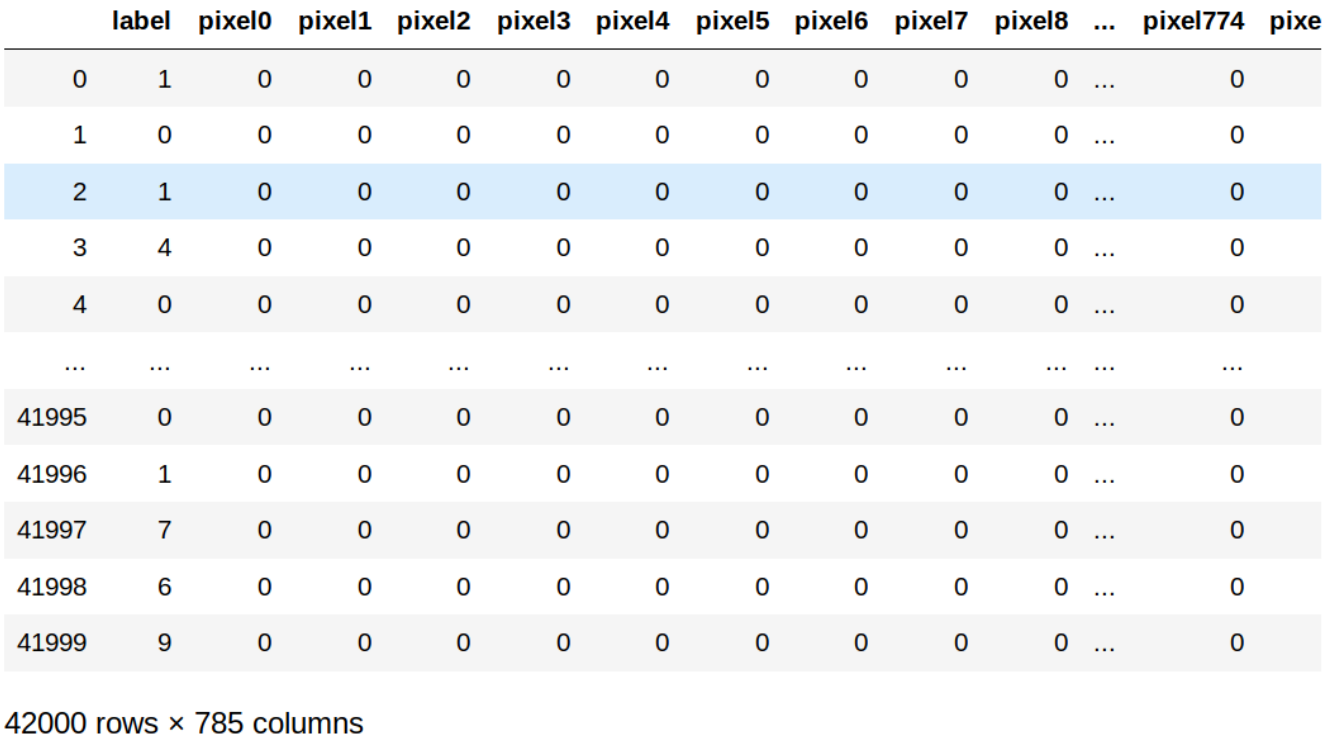


DIGIT CLASSIFIER

SAMPLE MNIST DATASET
(THE .CSV)

FEATURES:
PIXEL0 TO PIXEL783
(784 FEATURES)

TARGET:
LABEL
(0 TO 9)



	label	pixel0	pixel1	pixel2	pixel3	pixel4	pixel5	pixel6	pixel7	pixel8	...	pixel774	pixe
0	1	0	0	0	0	0	0	0	0	0	...	0	
1	0	0	0	0	0	0	0	0	0	0	...	0	
2	1	0	0	0	0	0	0	0	0	0	...	0	
3	4	0	0	0	0	0	0	0	0	0	...	0	
4	0	0	0	0	0	0	0	0	0	0	...	0	
...	...	...	...	...	...	...	...	...	...	...	...	...	
41995	0	0	0	0	0	0	0	0	0	0	...	0	
41996	1	0	0	0	0	0	0	0	0	0	...	0	
41997	7	0	0	0	0	0	0	0	0	0	...	0	
41998	6	0	0	0	0	0	0	0	0	0	...	0	
41999	9	0	0	0	0	0	0	0	0	0	...	0	

42000 rows × 785 columns

DIGIT CLASSIFIER

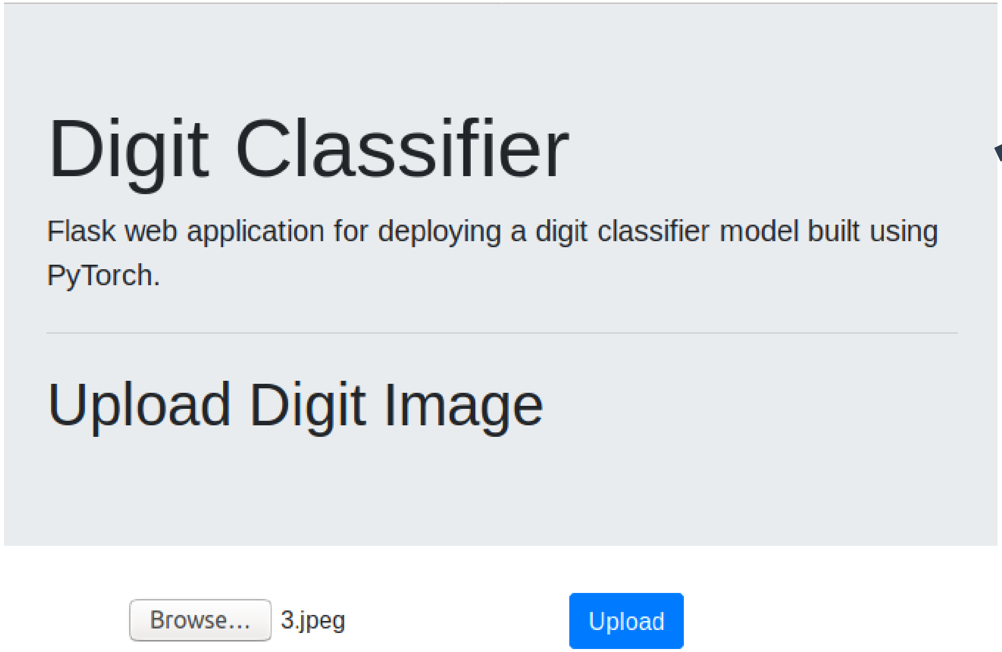
WHAT TOOLS WE’LL USE:

- PANDAS
- NUMPY
- MATPLOTLIB
- TENSORFLOW

CODE (INTRO TO ML/Digitclassifier/):

- MNIST using TensorFlow.html
- MNIST using TensorFlow.ipynb (NOTEBOOK)

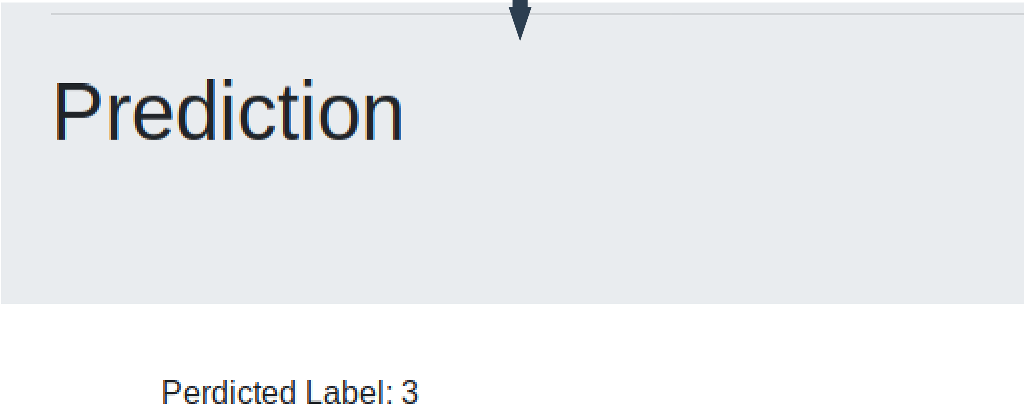
DIGIT CLASSIFIER



Digit Classifier

Flask web application for deploying a digit classifier model built using PyTorch.

Upload Digit Image



Prediction

Perdicted Label: 3

LINK: https://digit-class.herokuapp.com/

THE END
(THANK YOU)