

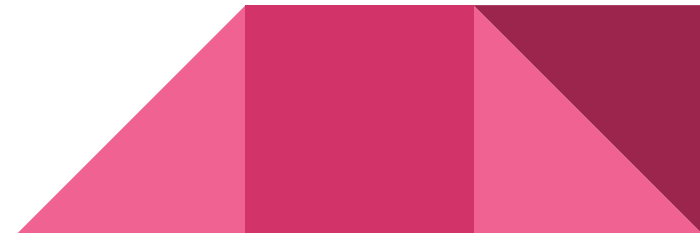


# Multiclass classification model for vote counting

David Lizama

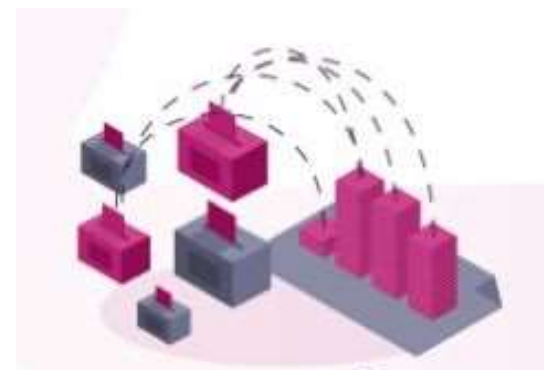
# Introduction

- I. The objective of this project is to detect the number of votes that each political party received in each section.
- II. The model can recognize handwritten numbers from a photo of the total votes of each polling place at the end of the elections.
- III. Recognition is useful for accelerating the process of counting votes. It would be possible to get a final result in only a few minutes.
- IV. A Gaussian Bayes classifier was used for this purpose.
- V. The train and test set had a good performance. Accuracy of 95%.



## Actual situation (Presidential elections Mexico 2018)

- ❑ Polling places are opened from 9 am to 6 pm.
- ❑ The count of votes takes between 1-2 hours.
- ❑ The 'Quick count' started at 10:30 and it released preliminary results using statistical methods.
- ❑ PREP lasts around 24 hours to count all votes.
- ❑ Federal Electoral Institute (INE) confirm final results at the end of the PREP.



# Project proposal: Handwriting recognition

Fill in total votes in a standard format poster.  
All numbers must be divided by squares.

Photo of the cartel should be sent by the president of each polling place.

A Machine Learning model recognize the number written by hand in order to count the votes received by each political party.

Results of each polling place are sent to the INE that counts votes to obtain the election results.



# Considerations

- ❑ All cartels should have a standard format in which all numbers should be divided by squares, and the order of political parties should be the same.
- ❑ Na's are preferred instead of zeros in the squares of the cartel.
- ❑ Photos should have a high resolution.

**IEEG**  
INSTITUTO ELECTORAL  
PROCESO ELECTORAL LOCAL 2020-2021  
RESULTADOS DE LA VOTACIÓN EN ESTA CASILLA  
6 DE JUNIO DE 2021

ENTIDAD FEDERATIVA: GUANAJUATO  
MUNICIPIO: Guanajuato  
DISTRITO ELECTORAL LOCAL: 0VIII  
SECCIÓN: 0849

ESCRIBA LOS RESULTADOS CON NÚMEROS GRANDES Y LEGIBLES.

PARTIDO COALICIÓN O CANDIDATURA	RESULTADOS (CON NÚMEROS)	
	DIPUTACIONES LOCALES	AYUNTAMIENTO
PAN	067	067
PR	017	017
PS	004	002
PR	004	003
PS	001	001
PR	012	013
PS	023	019
PR	002	003
PS	000	000
PR	003	004
PS	000	006
PR	000	000
PS	000	000
VOTOS PARA CANDIDATURAS NO REGISTRADAS	000	000
VOTOS NULOS	006	004

LA O EL PRESIDENTE DE LA MESA DIRECTIVA DE CASILLA  
ARCELA FERNÁNDEZ SAMPSON  
NOMBRE: \_\_\_\_\_ FIRMA: \_\_\_\_\_

**INE**  
Instituto Nacional Electoral

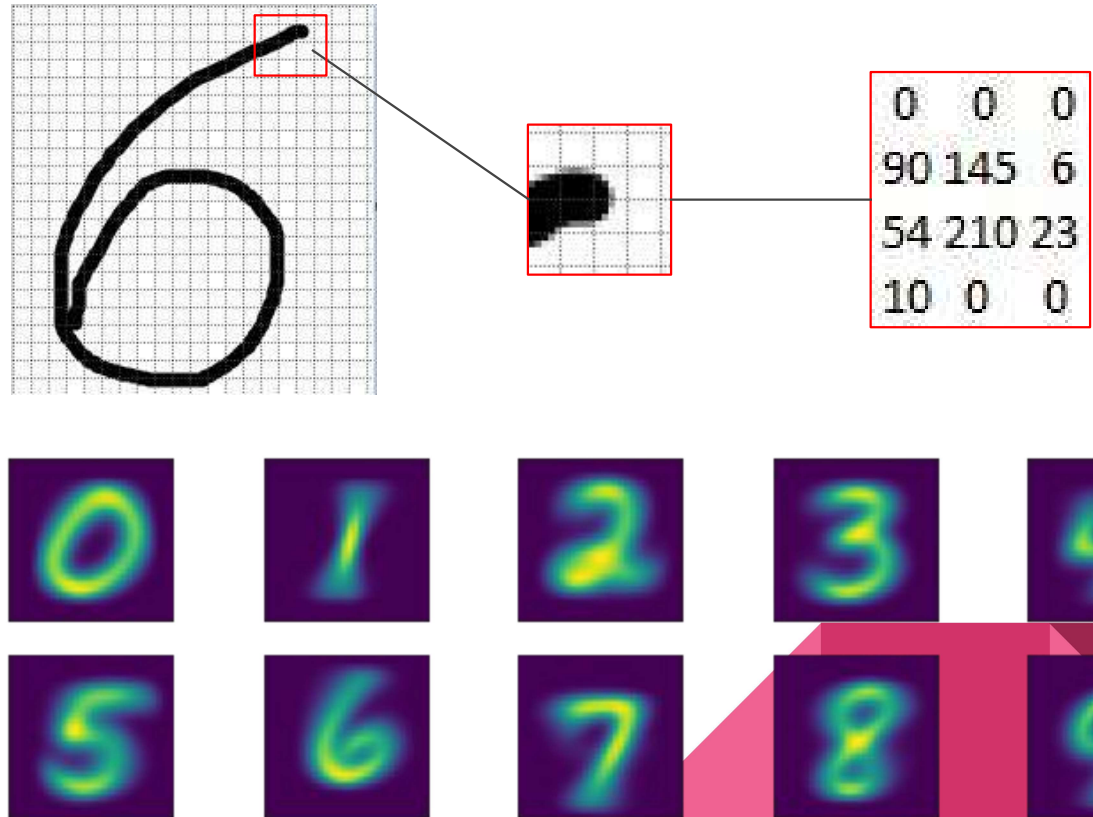
Estimada Ciudadanía:  
Este lugar ha pasado por un proceso de limpieza y  
desinfección posterior a la Jornada Electoral  
Gracias por participar

ELECCIONES 2021

# Data

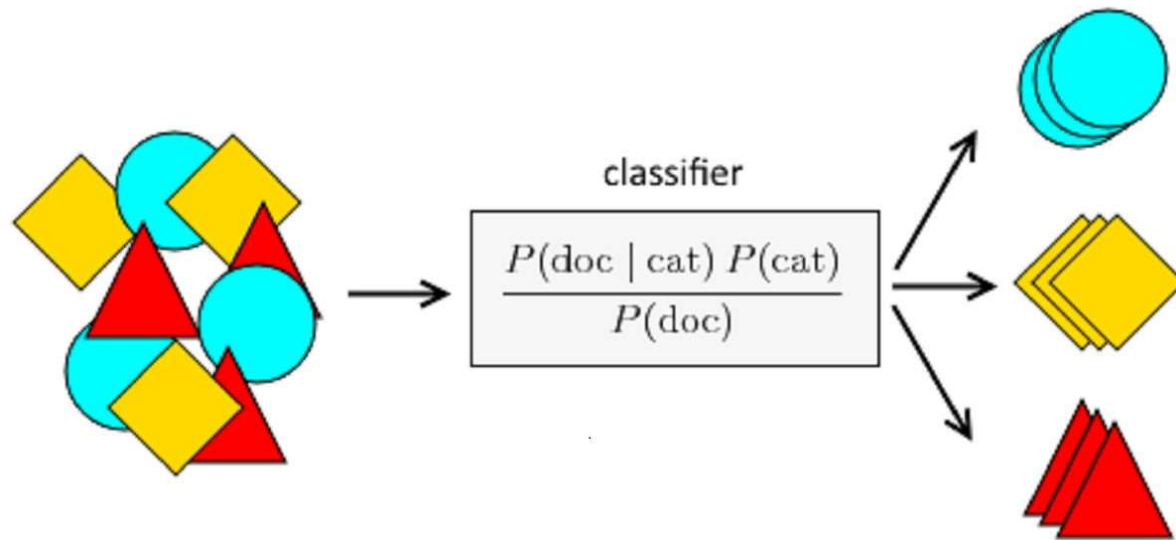
A set of 60,000 images are split into two data sets: train (50,000) and test (10,000).

Each observation has an array of 784 numbers (square matrix dim 28), which indicate the percentage of color in a pixel. The range is from 0 to 255.



# Model

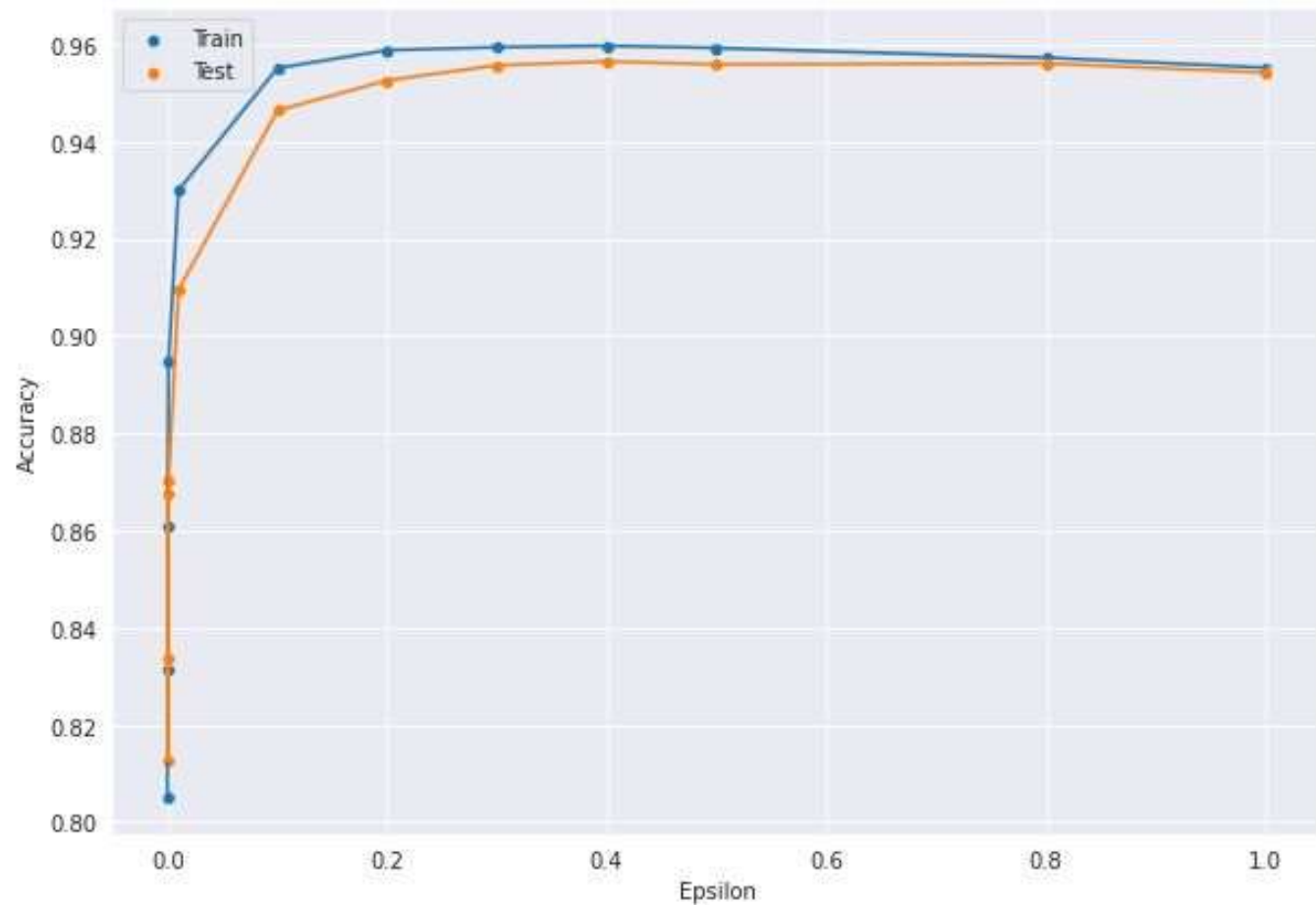
- ❑ Gaussian Naive Bayes model was used to solve this problem.
- ❑ The performance of the model were improved by training different models with different tuning parameters.
- ❑ Preprocessing: standardization of test and train.





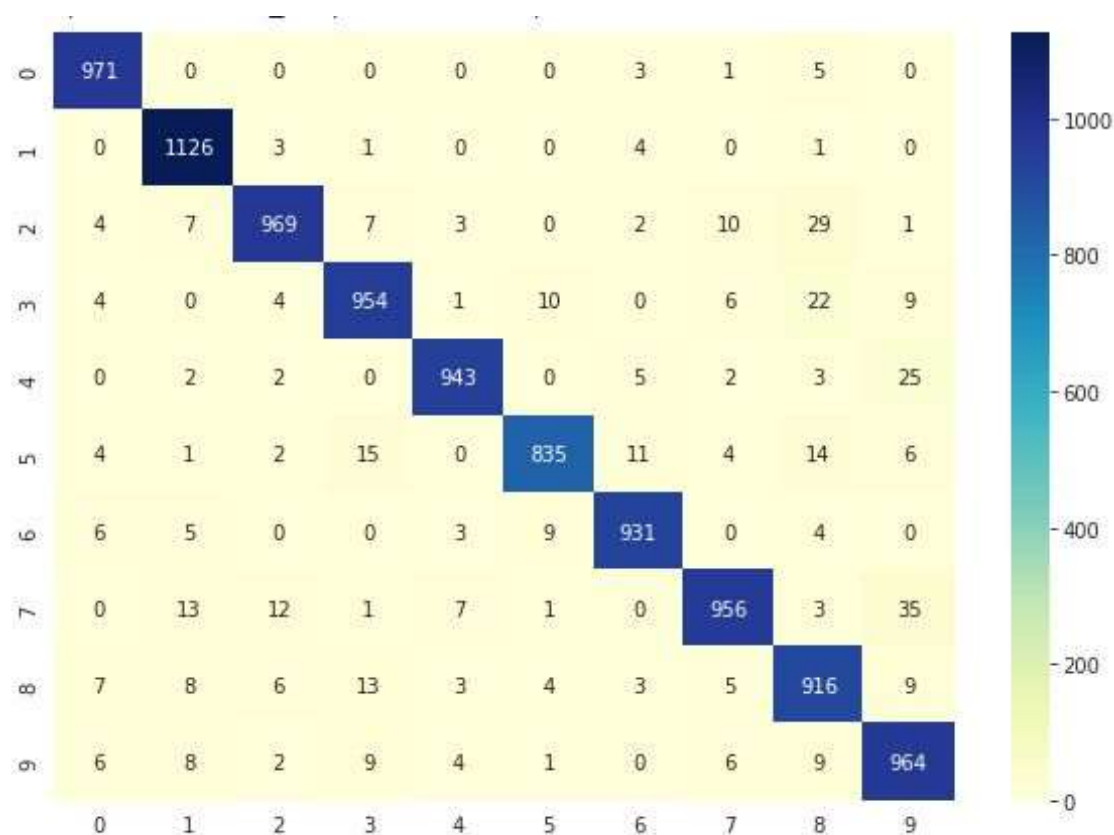
# Performance: Train and Test Set

	Epsilon	Train	Test
0	0.000001	0.805333	0.8676
1	0.000010	0.831517	0.8125
2	0.000100	0.860767	0.8335
3	0.001000	0.894983	0.8703
4	0.010000	0.929967	0.9095
5	0.800000	0.957283	0.9561
6	0.500000	0.959283	0.9559
7	0.400000	0.959717	0.9565
8	0.300000	0.959433	0.9557
9	0.200000	0.958817	0.9526
10	0.100000	0.955117	0.9464
11	0.500000	0.959283	0.9559





## Performance: Confusion Matrix of test set



- ❑ We got an accuracy of 95% in both sets.
- ❑ Main errors are found in the detection of number 7 (prediction 0) and number 2 (prediction 5).

# Conclusions

- ❑ The proposal project is taking photos of each polling place and processing the information via handwriting recognition.
- ❑ Handwriting recognition could accelerate the process of counting votes with high precision.
- ❑ Instead of waiting 24 hour for final results, INE will be able to count votes in just a few minutes.
- ❑ Performance in both data sets were quite similar. Accuracy of 95%.
- ❑ Bad records should not affect the final results.

