



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ  
UNIVERSITY OF PIRAEUS

DEPARTMENT OF DIGITAL SYSTEMS



NCSR DEMOKRITOS  
INSTITUTE OF INFORMATICS AND  
TELECOMMUNICATIONS

# Machine Learning Project

## Covid-19 Mortality Prediction

Gkatsis Vasileios

MTN 2004

# Overview

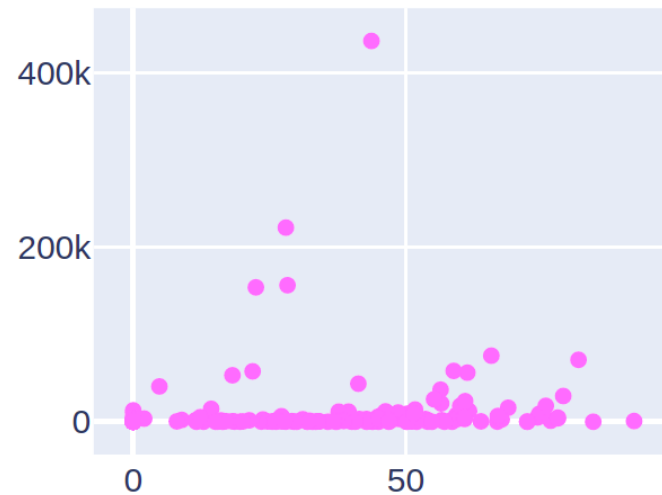
Create machine learning models capable of predicting the number of people dying from covid-19 in a country during a month.

- Gather and process data. ([Our World In Data](#))
- Extract the more explanatory features.
- Create and tune models.
- Evaluate the models.

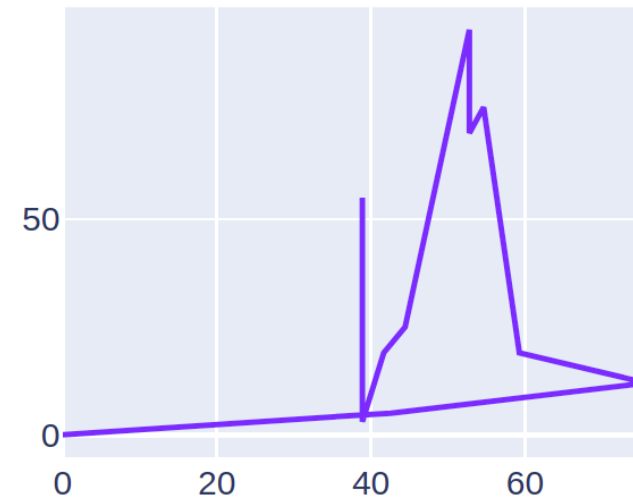
# Feature Extraction

Choose the features that explain better the data using simple graphs.

Total Deaths - Smokers



Total Deaths – Stringency Index



# Create Models

- Linear Regression
- Lasso Regression
- Ridge Regression
- K-NN Regression
- Elastic-Net
- Support Vector Machines

(SGD, Decision Trees and Tweedie Regression  
Where also tested but left out due to very bad  
results)

Python library: scikit-learn

# Tune Models

Split data on train and validation.

Perform k-fold on train data and find average MAE.

Find MAE of validation dataset.

Keep the combination of parameter values that minimize both MAE scores

# Test

Make predictions for the test dataset.

Count performance using the following metrics:

- $R^2$
- Mean Average Error (MAE)
- Mean Squared Error (MSE)
- Rooted Mean Squared Error (RMSE)

# Results

	Linear Regression	Lasso Regression	Ridge Regression	kNN Regression	Elastic Net	SVM
R2	0.79	0.78	0.69	0.2	0.78	-0.15
MAE	2326.4	2360	2438.9	4643.7	2419.1	5232.1
MSE	18312097	19008276	27148578	69774062.2	19427773	100242603
RMSE	4279.26	4359.8	5210	8353	4407.7	10012

# Final Thoughts



- Not a feasible project.
- Not with simple machine learning tools.
- Not without including more data (epidemiological, cultural etc.)
- Maybe deep learning methods can produce more stable models.