CLASSIFICACIÓ APC

JORDI GONZALEZ

ALEX GALVANY

DANIEL ALCOVER

OBJECTIUS DE LA PRÀCTICA

Aplicar diferents models classificadors

- SVM
- Regressor logístic

Entendre les millores d'aplicar kernels

Avaluar correctament l'error del model

Visualitzar les dades juntament amb el model Ser capaç d'aplicar tècniques de classificació en casos reals i validar el resultats.



pH Value

Hardness

Solids

Chloramines

Sulfate

Conductivity

Organic Carbon

Trihalomethanes

Turbidity

Potability



ph -	1	0.082	-0.089	-0.034	0.018	0.019	0.044	0.0034	-0.039	-0.0036
Hardness -	0.082	1	-0.047	-0.03	-0.11	-0.024	0.0036	-0.013	-0.014	-0.014
Solids -	-0.089	-0.047	1	-0.07	-0.17	0.014	0.01	-0.0091	0.02	0.034
Chloramines -	-0.034	-0.03	-0.07	1	0.027	-0.02	-0.013	0.017	0.0024	0.024
Sulfate -	0.018	-0.11	-0.17	0.027	1	-0.016	0.031	-0.03	-0.011	-0.024
Conductivity -	0.019	-0.024	0.014	-0.02	-0.016	1	0.021	0.0013	0.0058	-0.0081
Organic_carbon -	0.044	0.0036	0.01	-0.013	0.031	0.021	1	-0.013	-0.027	-0.03
Trihalomethanes –	0.0034	-0.013	-0.0091	0.017	-0.03	0.0013	-0.013	1	-0.022	0.0071
Turbidity -	-0.039	-0.014	0.02	0.0024	-0.011	0.0058	-0.027	-0.022	1	0.0016
Potability -	-0.0036	-0.014	0.034	0.024	-0.024	-0.0081	-0.03	0.0071	0.0016	1
	- yd	lardness -	Solids -	ramines -	Sulfate -	ductivity -	c_carbon -	rethanes -	Turbidity -	otability -

- 0.8

- 0.6

- 0.2

- 0.0

MODEL SELECTION

Model	Accuracy (%)			
KNN	64,02%			
SVM	68,14%			
Random Forest	69,51%			
Decision Tree	64,02%			
Logistic Regression	62,80%			

CROSS-VALIDATION

Evitar el overfitting

$$K = 5$$

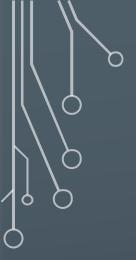
LeaveOneOut?

METRIC ANALYSIS

Accuracy Score

F1_Score

Average Precision Score

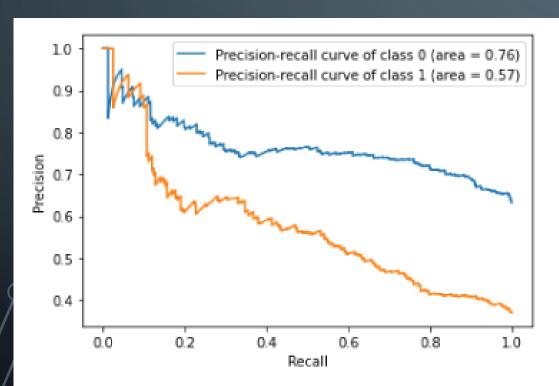


HYPERPARAMETER SEARCH

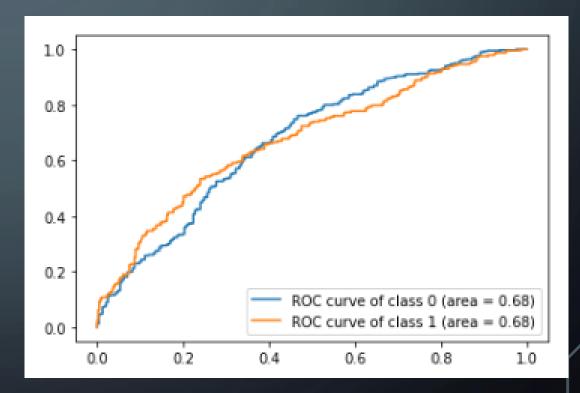
- Provar motls paràmetres d'entrada
- Alt cost computacional
- Amb recursos limitats escolliríem el mètode de:
 - Randomized Parameter Optimization

COMPARATIVA DE MODELS

Precission-Recall Curve



ROC Curve



CONCLUSIONS