


PRÀCTICA 2

# **CLASSIFICACIÓ DATASET TITANIC**

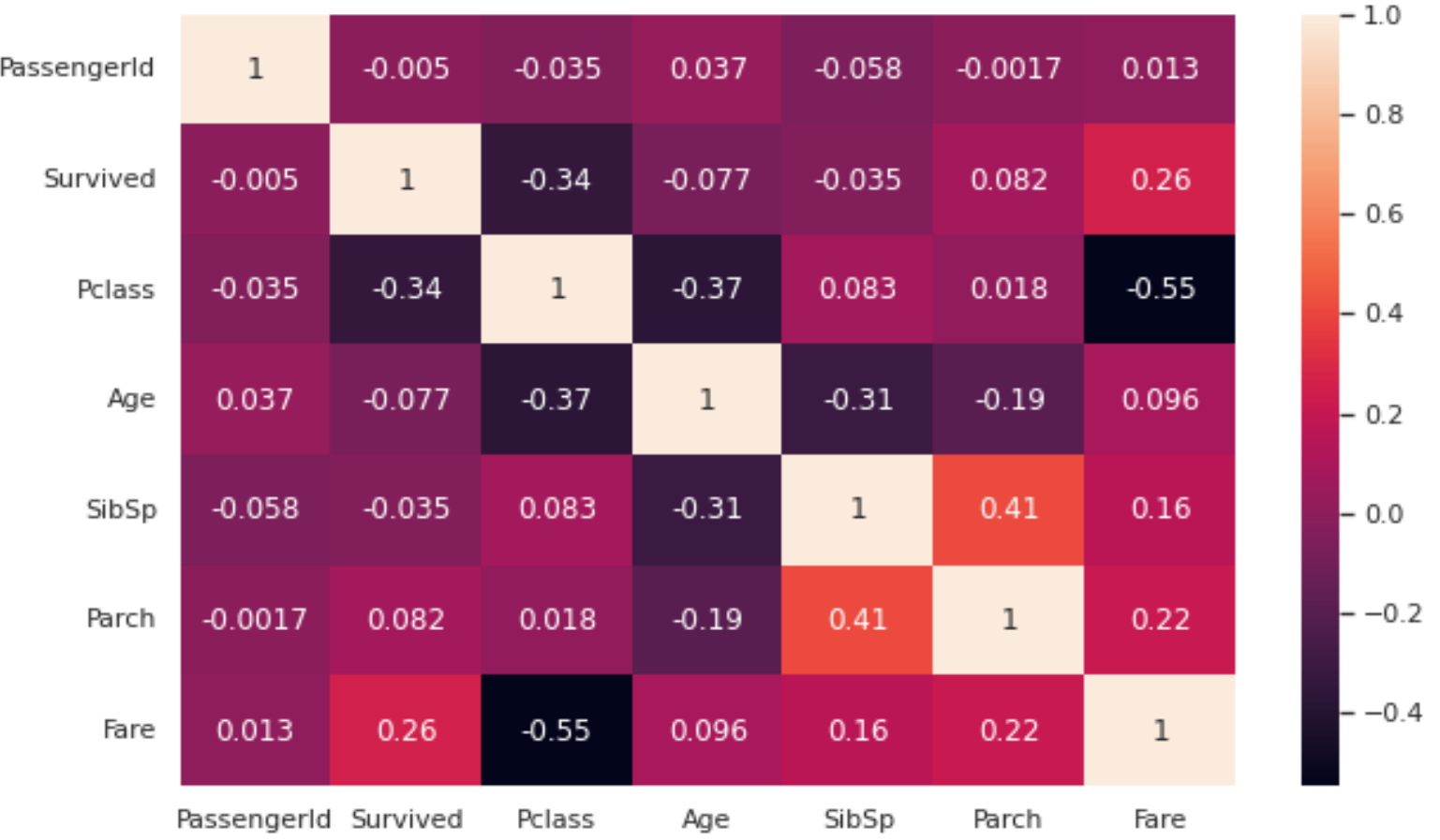
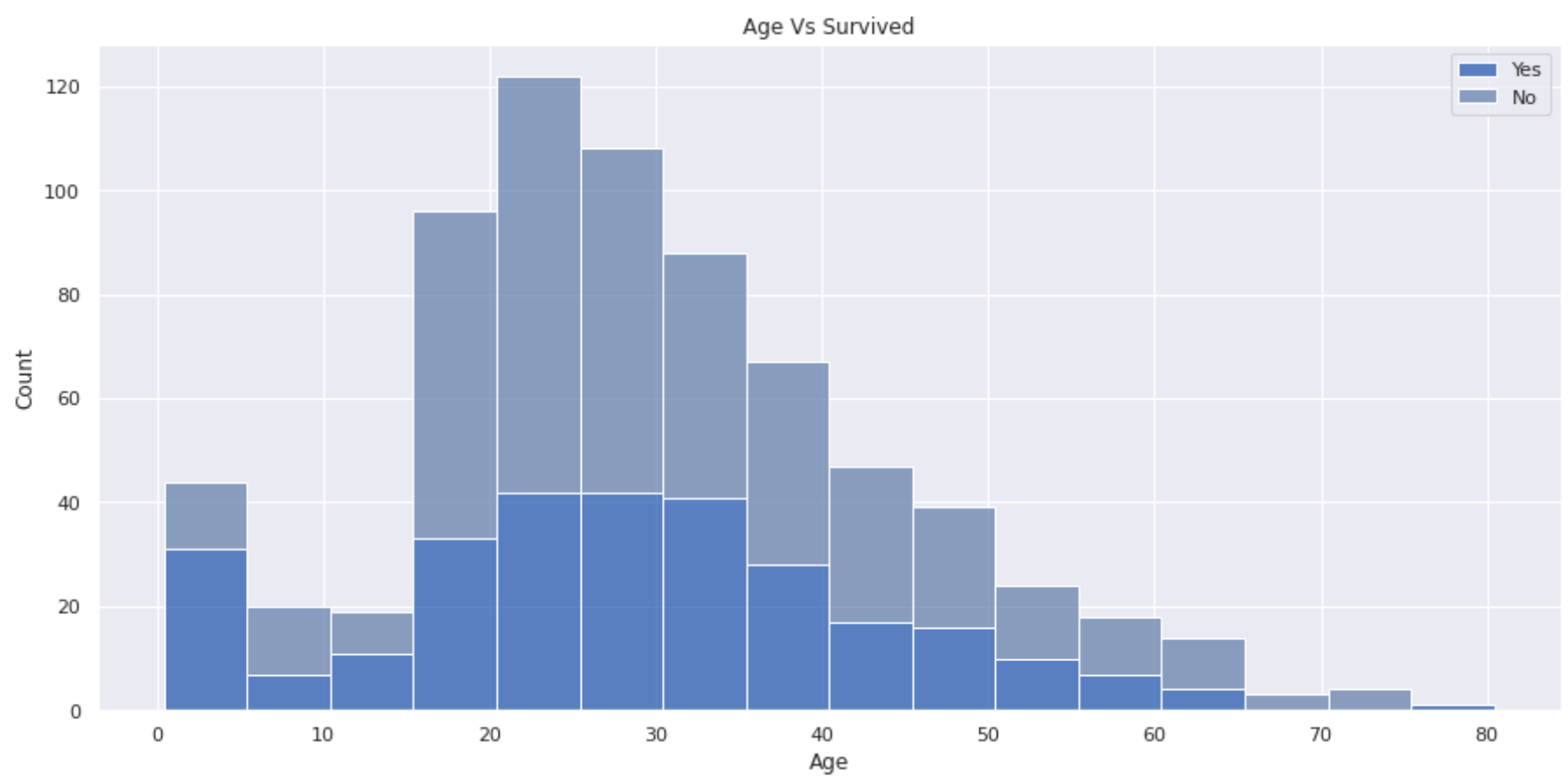
BIEL GONZÀLEZ  
JUDIT PANISELLO  
CRISTINA SOLER

- 
- 01 INTRODUCCIÓ**
  - 02 EDA**
  - 03 PROCESSAMENT DE DADES**
  - 04 MODELS**
  - 05 MILLOR MODEL**
  - 06 APARTAT A**
  - 07 CONCLUSIONS**

# INTRODUCCIÓ



# EDA



# PROCESSAMENT DE DADES

**1**

**TRACTAMENT DE  
NULLS**

**2**

**ELIMINACIÓ DE  
VARIABLES**

**3**

**ONE HOT**

**4**

**PCA DEL MODEL**

# MODELS

## Regressió logística

- Accuracy: 83.11%

## Perceptró

- Accuracy: 74.67%

## KNN

- Accuracy: 82.84%

## SVC

- **SVC general:**
  - Accuracy: 83.11%
- **Lineal:**
  - Accuracy: 83.11%

## Ensemble

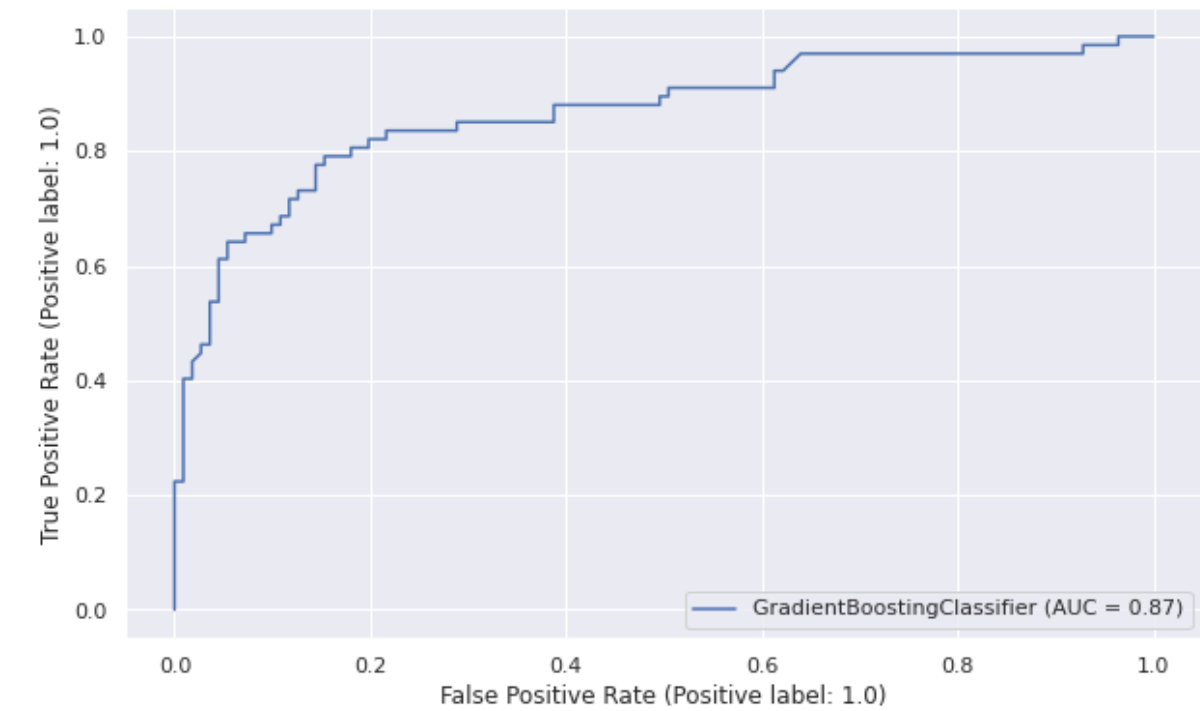
- **Ada boosting**
  - Accuracy: 83.95%
- **Random Forest**
  - Accuracy: 83.68%
- **Gradient Boosting**
  - Accuracy: 83.96%

# MODELS - TEMPS D'EXECUCIÓ (COLAB)

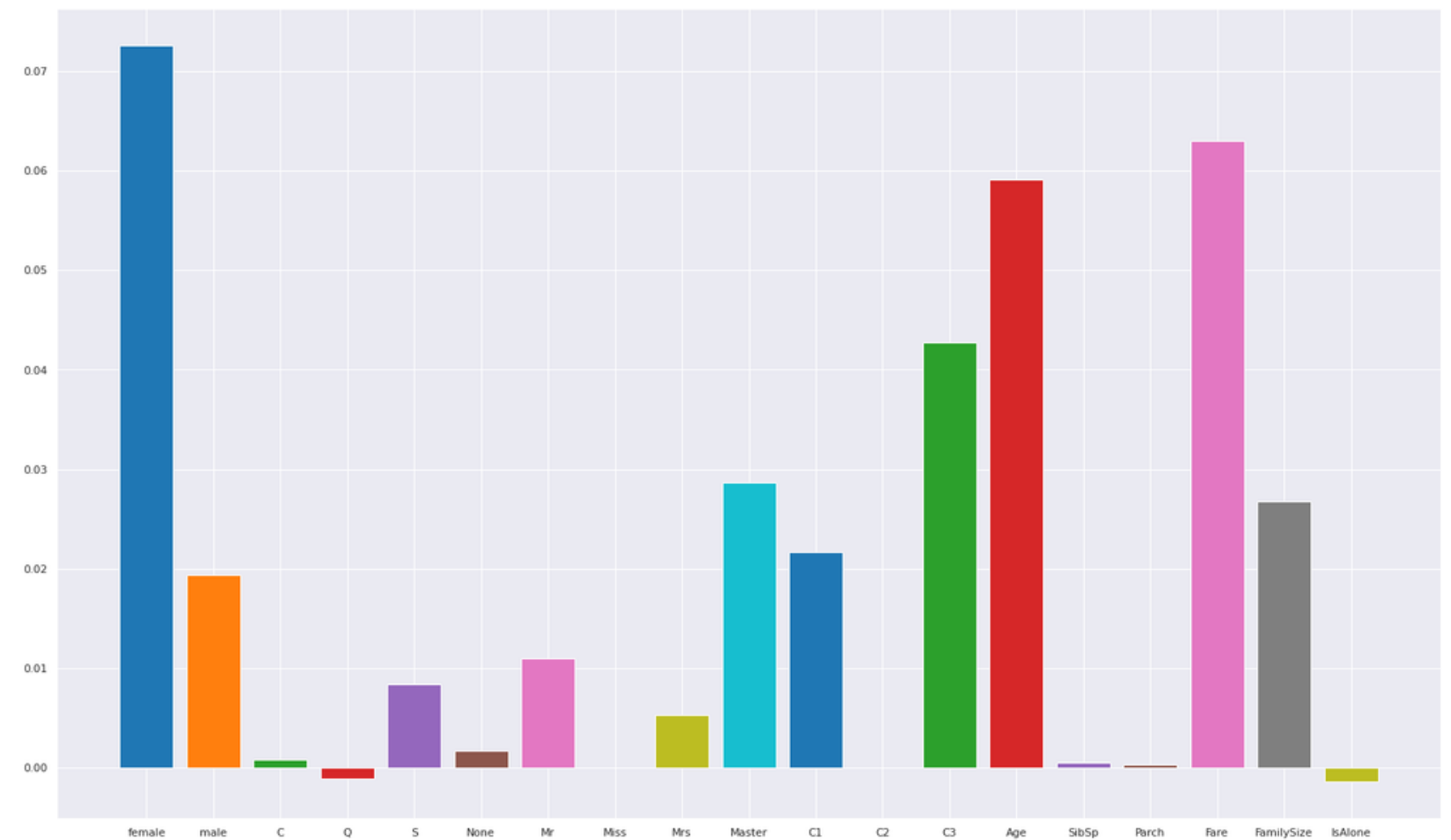
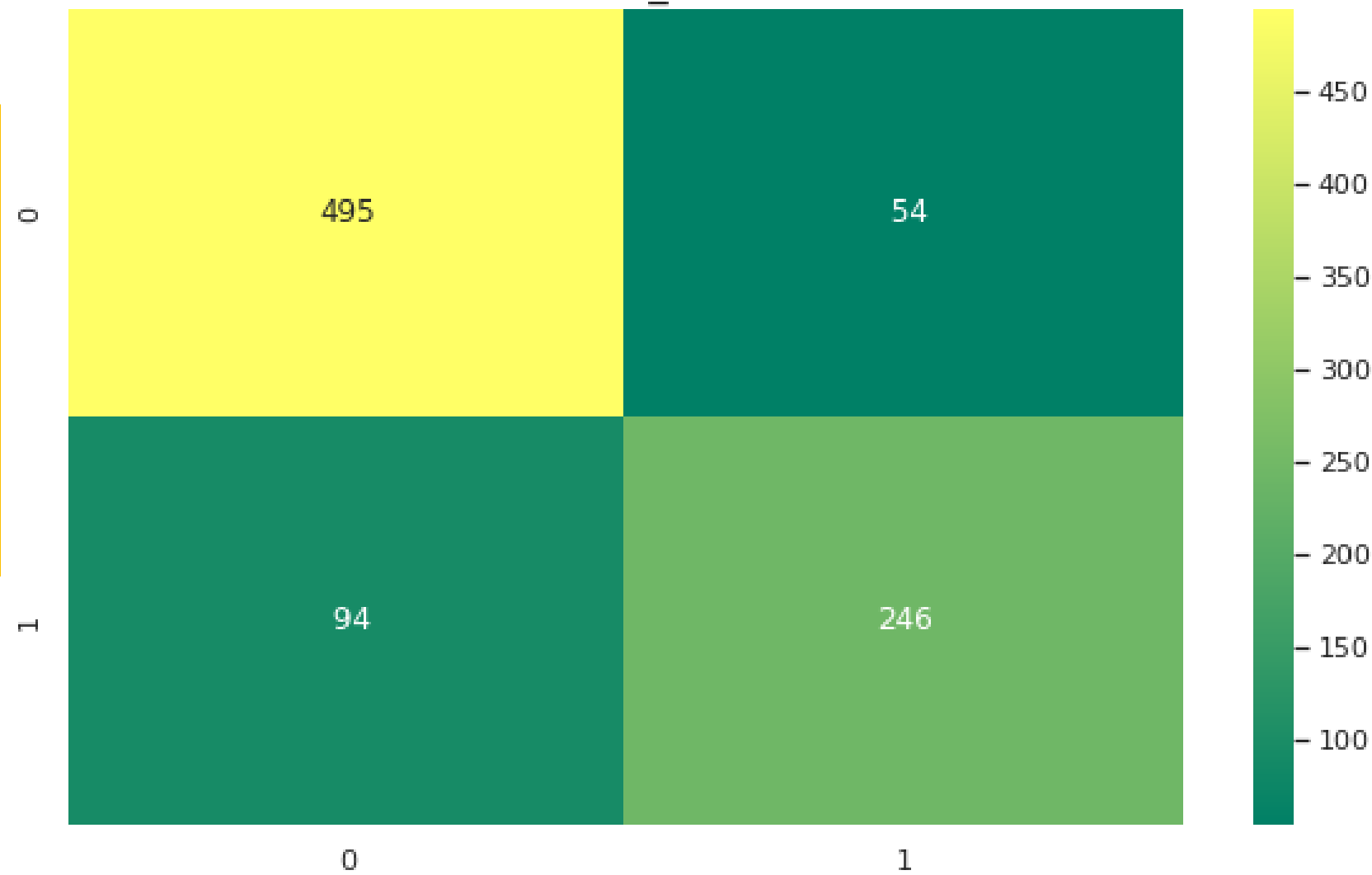
Models	Time (seconds)	
	Logistic Regression	0.948
	Logistic Regression (LOOCV)	42.751
	Perceptron	0.710
	LinearSVC	1.061
	SVC	2.624
	KNeighborsClassifier	255.889
	AdaBoosting + Logistic	903.213
	RandomForestClassifier	154.613
	GBClassifier	27.243

# MILLOR MODEL

El millor model que hem obtingut és el Gradient Boosting.



Confusion\_matrix





# APARTAT A

		Grid Search Results					
Models	DecisionTreeClassifier	0.816	0.163	0.870	0.837	0.836	0.836
	SVC	0.838	0.169	0.834	0.831	0.830	0.830
	RandomForestClassifier	0.837	0.174	0.867	0.826	0.824	0.824
	LogisticRegression	0.835	0.191	0.870	0.809	0.809	0.809
	KNeighborsClassifier	0.828	0.157	0.876	0.843	0.844	0.839
	GBClassifier	0.838	0.157	0.859	0.843	0.842	0.841
		Accuracy	MSE	ROC	Recall	Precision	F1-Score

## CONCLUSIONS

Després de provar amb diferents mètriques i millors paràmetres per classificadors obtenim que el Gradient Boosting ens dona els millors resultats. Per temps d'execució també és dels més ràpids.

El random forest també es podria considerar com bo.