

Cas **Kaggle** GPU Kernel Performance



Cristina Soler
NIU: 1603542

<https://github.com/kermitsc7/GPU-Kernel-Performance->



Índex

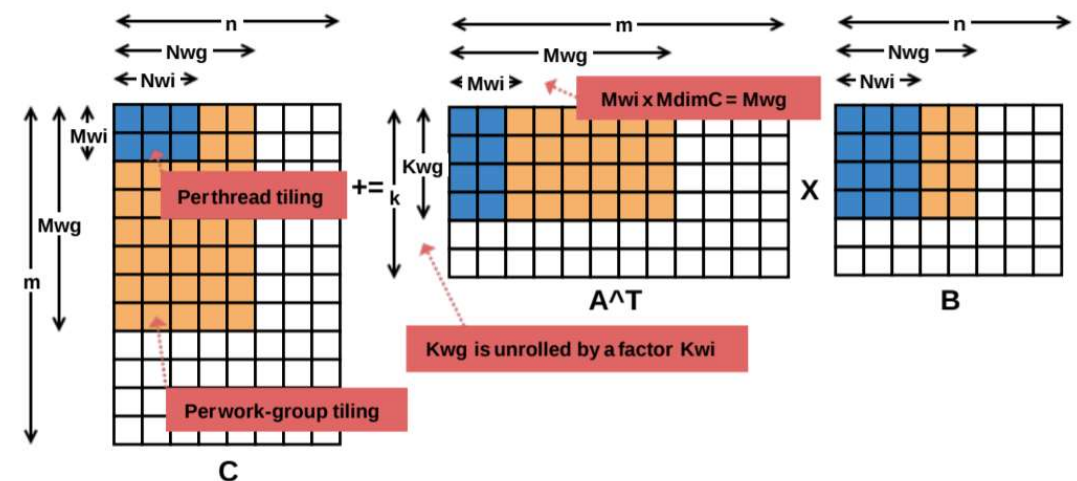
- Introducció i Objectius
- Descripció de les dades
- Processament i tractament de dades
- Avaluació dels models
 - Regressió
 - Classificació
- Dashboard



Introducció i Objectius

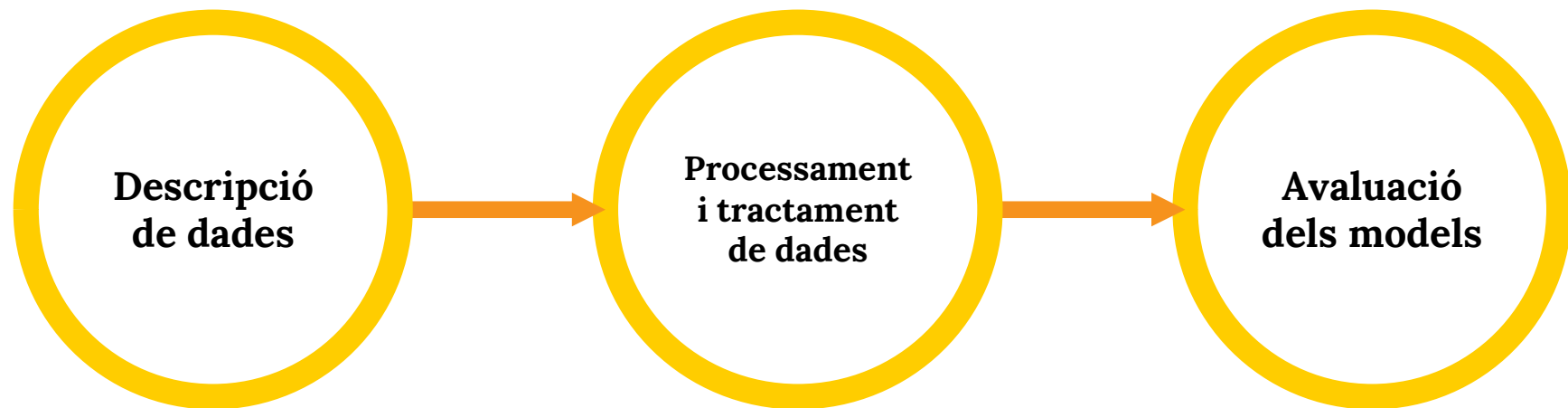
Aquest conjunt de dades mesura el temps d'execució d'un producte matriu-matriu $A \cdot B = C$, on totes les matrius tenen una mida de 2048 x 2048.

Objectiu: predir el temps d'execució segons els paràmetres del processador





Passos realitzats



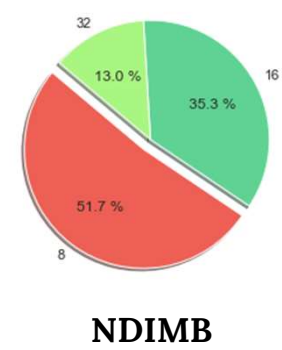
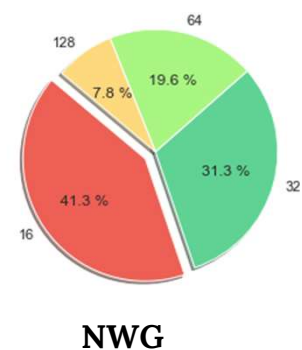
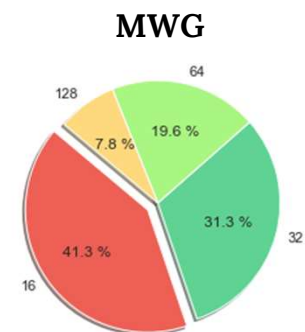
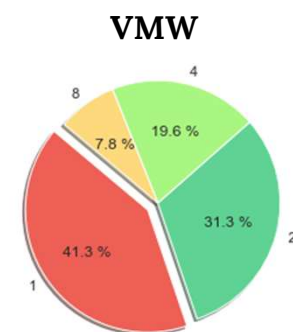
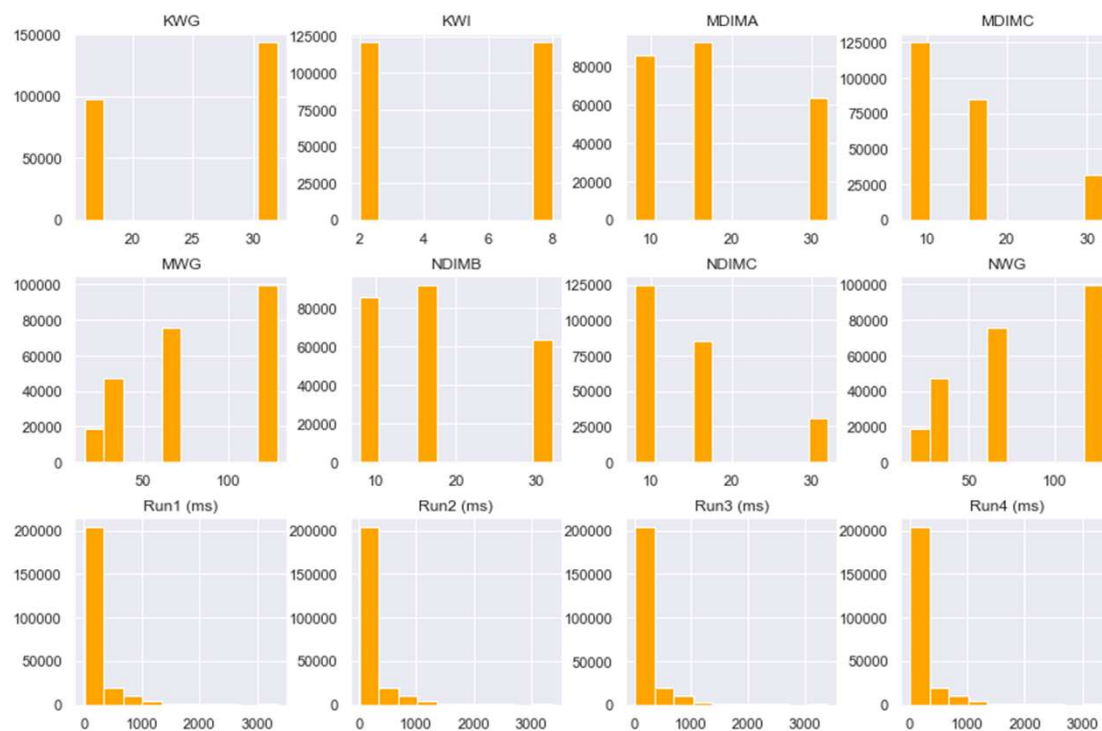


Descripció de les dades

- **Les variables independents:**
 - 10 variables de nombres enters (*int*)
 - 4 variables categòriques
- **Variable Objectiu:**
 - 4 execucions de temps diferents usant els mateixos paràmetres (*float*)
- **No tenim nuls**



Descripció de les dades





Processament i tractament de dades

CREACIÓ DE LA VARIABLE OBJECTIU

Realitzem la mitjana de les dades de sortida, per crear una variable nova

C

N

NOVES VARIABLES

A partir de la informació que tenim creem noves variables

Apliquem un one-hot encoder a les variables categòriques

ONE-HOT ENCODER

O

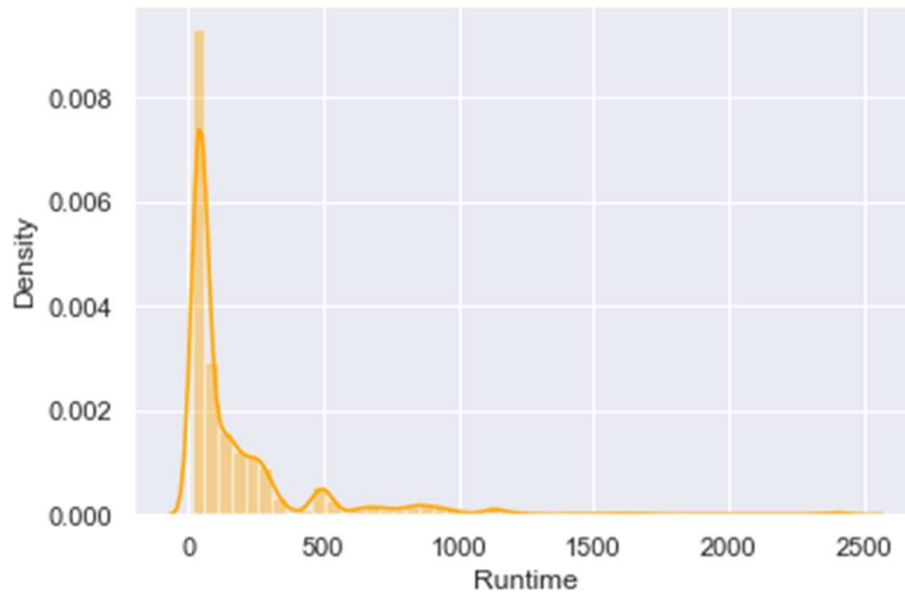
E

Eliminem els outliers a partir del mètode Z-score

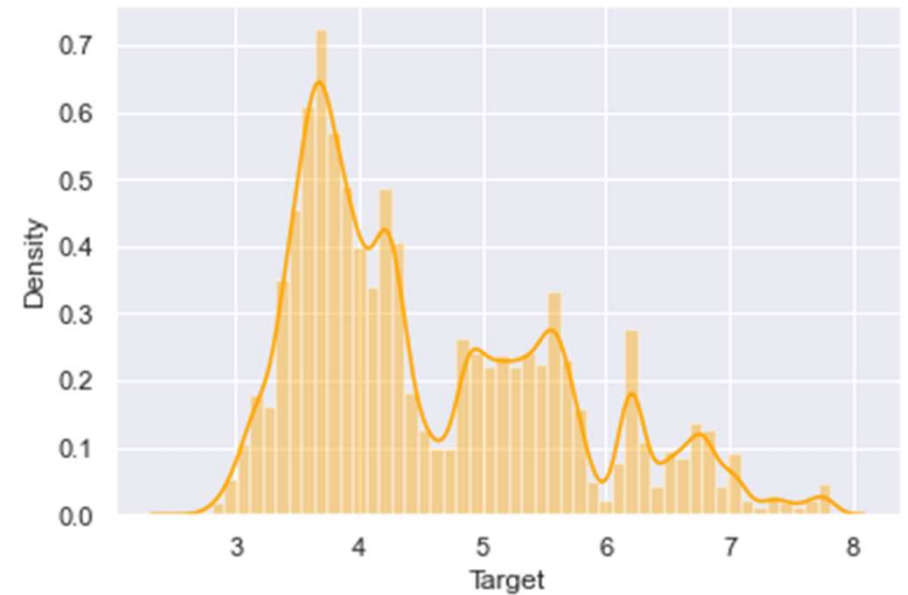
ELIMINACIÓ OUTLIERS



Transformació de la variable objectiu

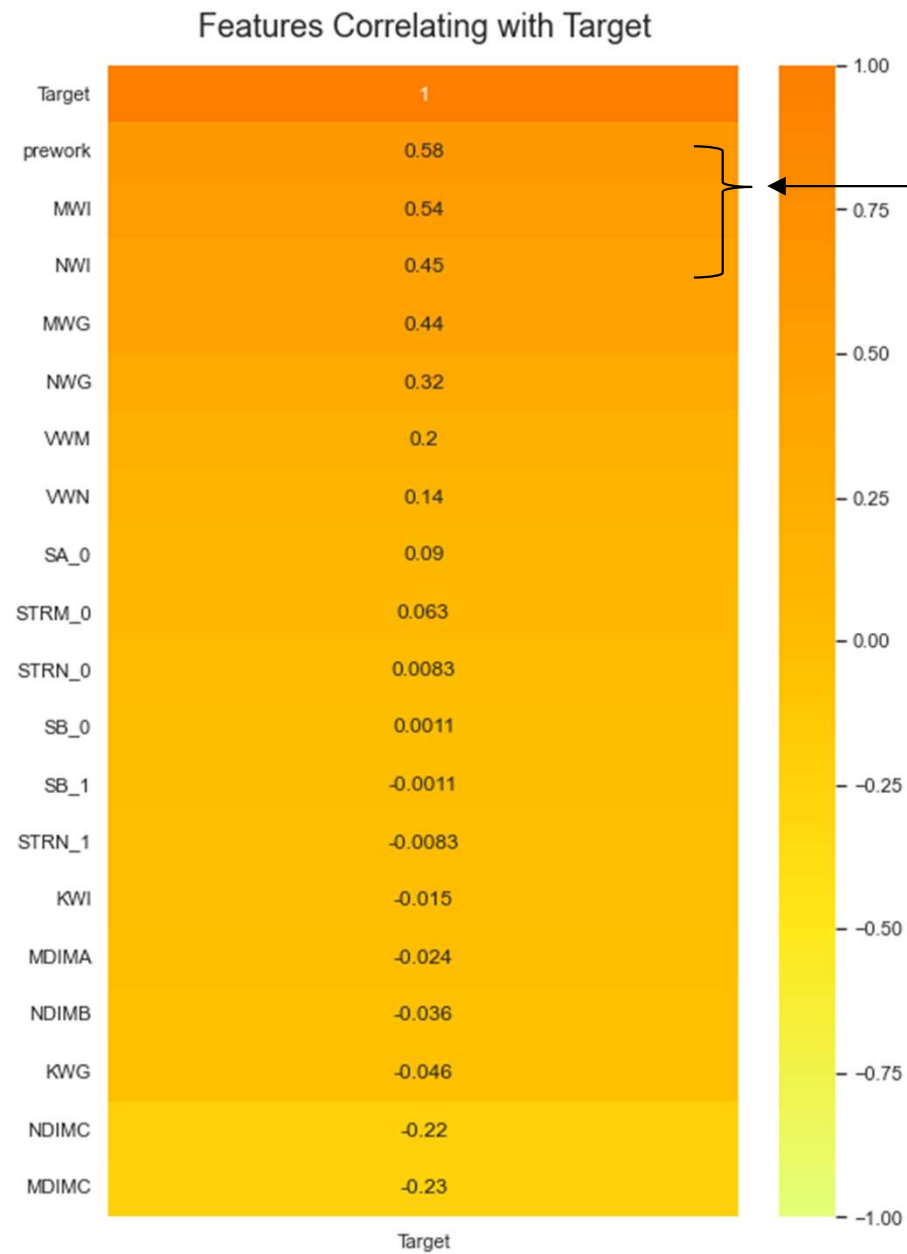


Abans de la transformació



Després de la transformació

Correlació de les dades al finalitzar el processament :



Les variables creades són les de millor correlació

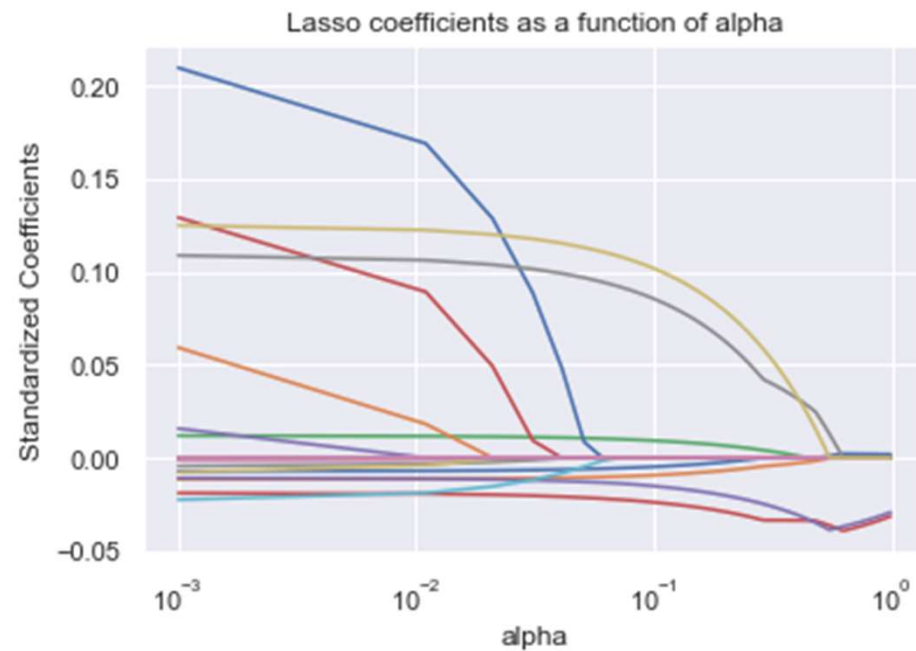


Models

- Regressió
- Classificació

Lasso Feature Selection

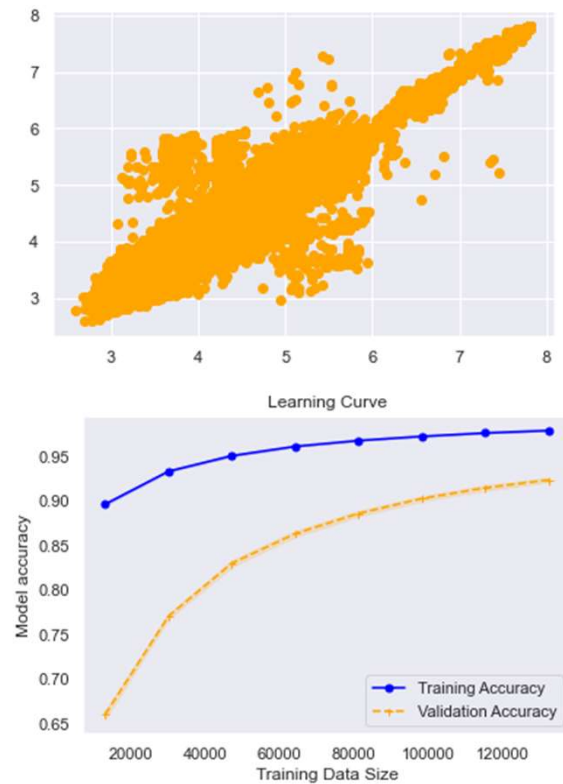
- A partir de la regressió **Lasso** buscarem quins són aquells coeficients amb més importància al model, d'aquesta manera reduïrem atributs amb els que treballarem.
- Arribem a $\alpha = 0,001$ on 4 predictors tenen coeficient 0



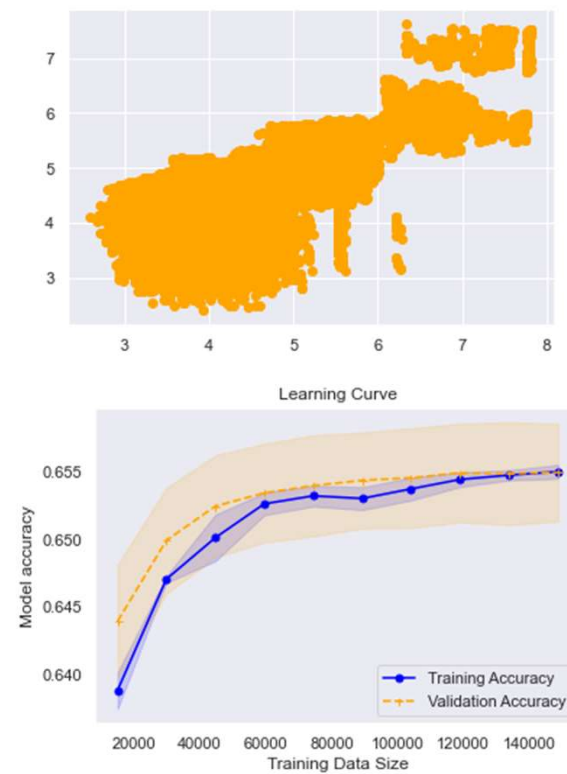


Regressió

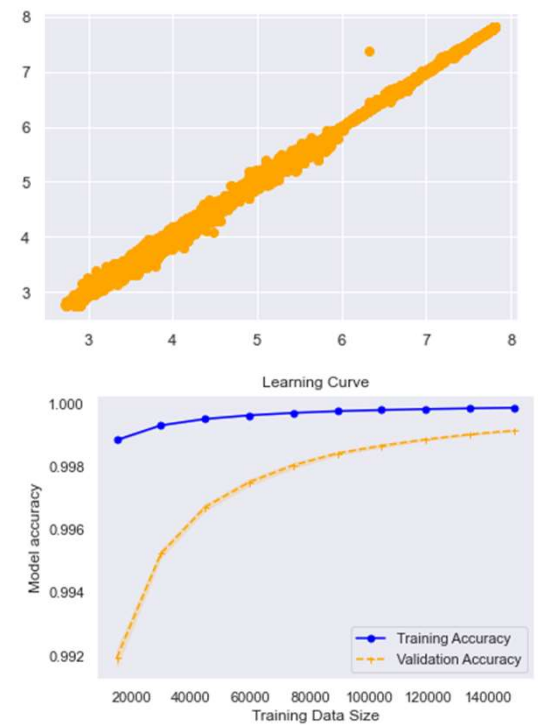
● KNN Regressor



● SGD Regressor



● RF Regressor





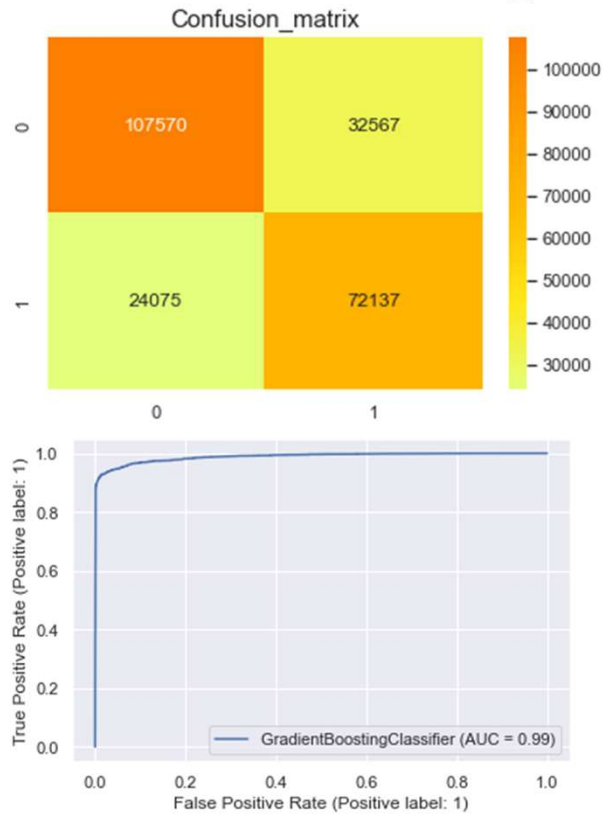
Resultats Regressió

		Comparació de models			
Models	KNNRegressor	0.93374	0.07859	0.28034	15.10953
	SDGRegressor	0.65534	0.40880	0.63937	0.47100
	RandomForestRegressor	0.99926	0.00088	0.02960	57.81041
		R2 score	MSE	RMSE	Time

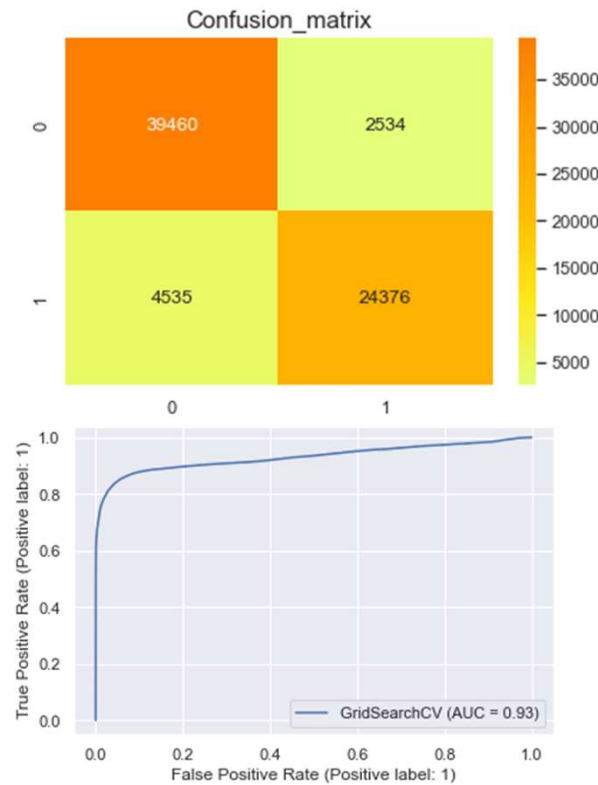


Classificació

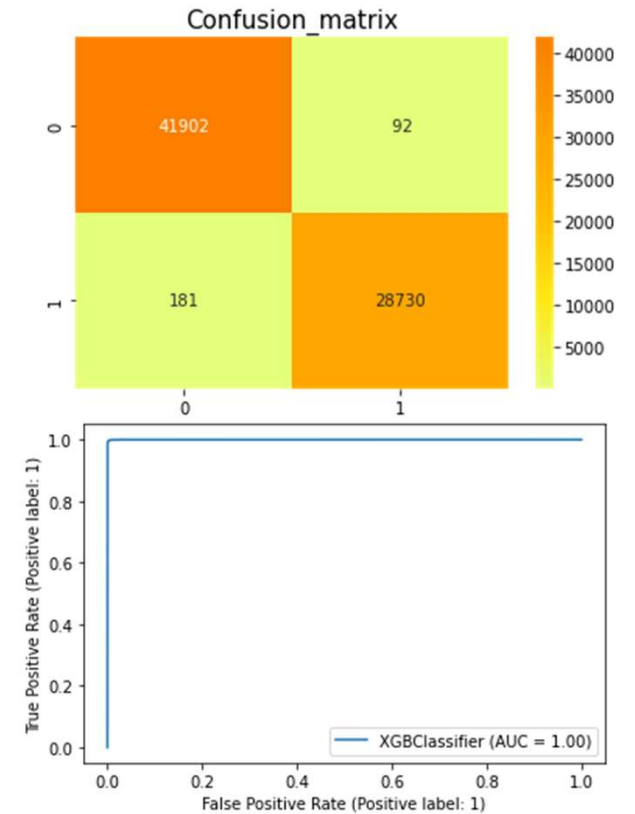
Gradient Boosting



Regressió Logística



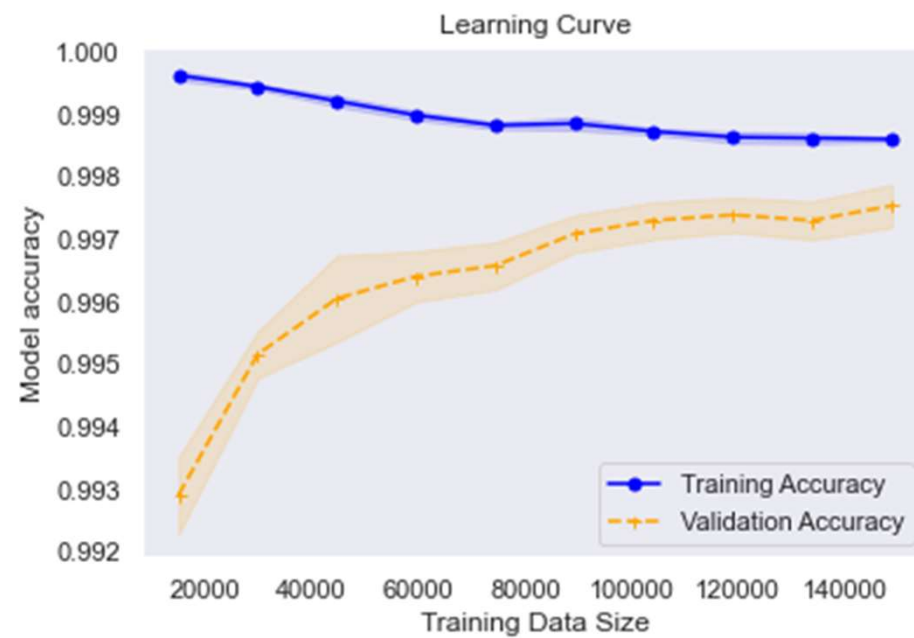
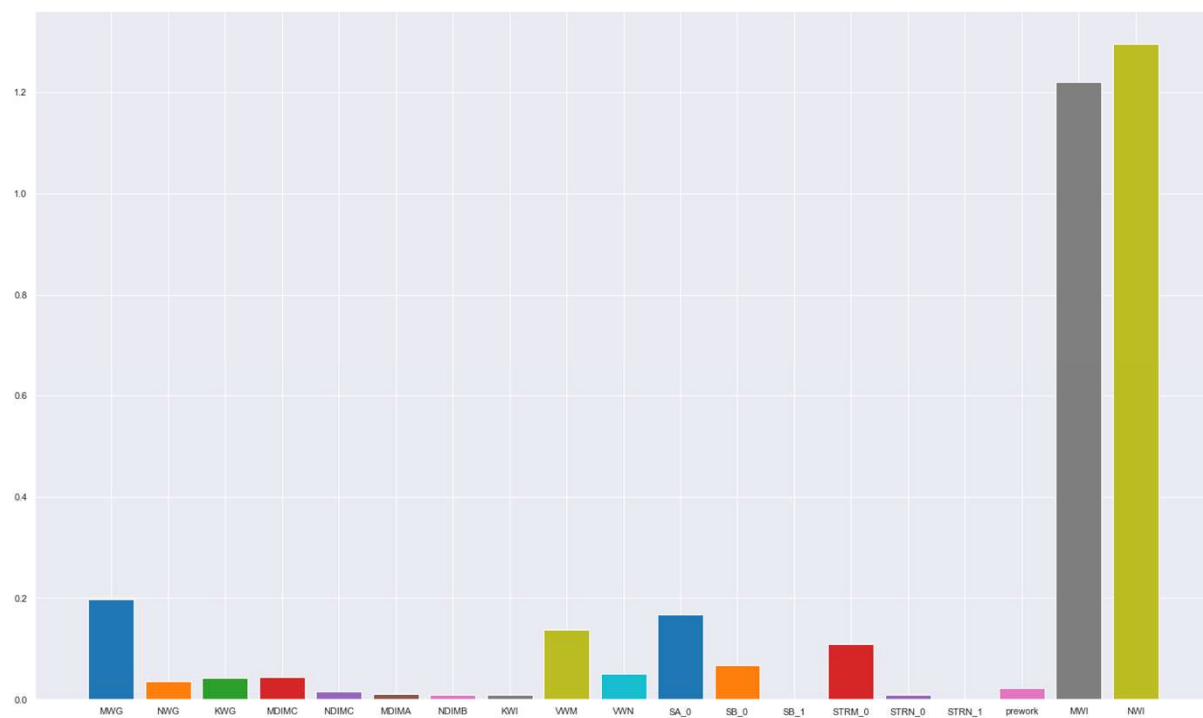
XGboost





Classificació

XGboost





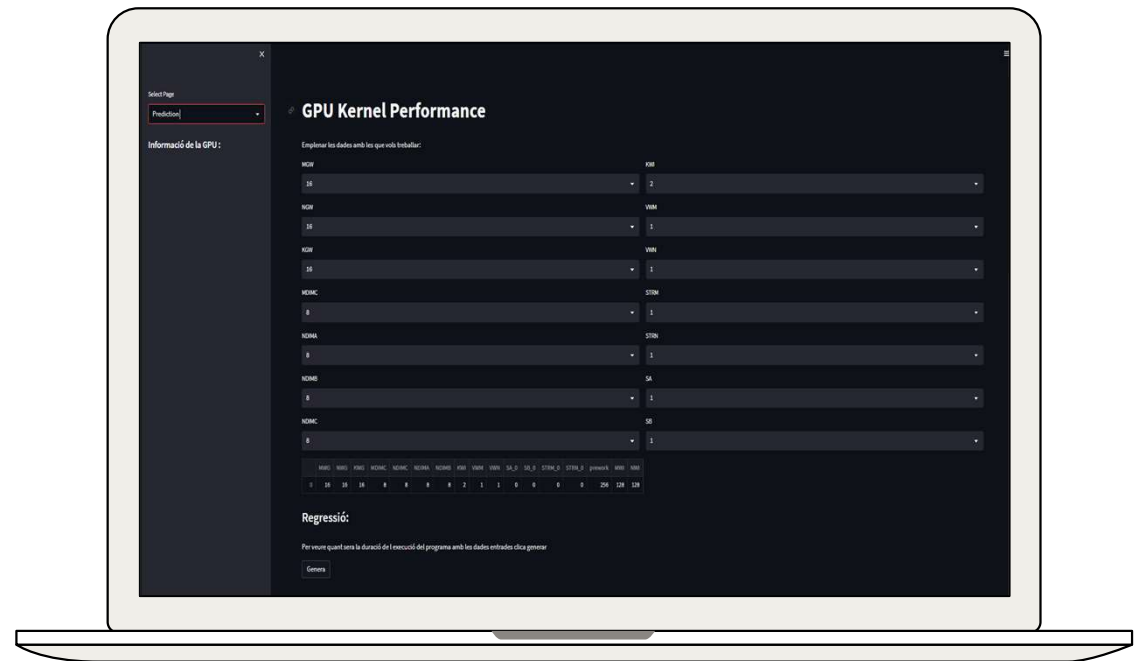
Resultats Classificació

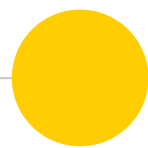
		Grid Search Results						
Models	LogisticRegression	0.900	0.101	0.927	0.899	0.900	0.899	20.82323
	XGBoost	1.000	0.002	1.000	0.998	0.998	0.998	4.29226
	GBClassifier	0.900	0.033	0.995	0.967	0.968	0.967	2323.81650
		Accuracy	MSE	ROC	Recall	Precision	F1-Score	Time



Dashboard / Demo

Ens mostra l'anàlisi completa de les dades, i ens permet realitzar prediccions a partir de les dades que li entrem.





18