



EMAIL SPAM CLASSIFICATION

Nom: Arnau Cruz Gargallo

Link Github: <https://github.com/arnaucruz1997/APC-Cas-Kaggle>

Introducció

Mostres: 5172

Atributs: 3002

Objectiu:

NO SPAM



SPAM



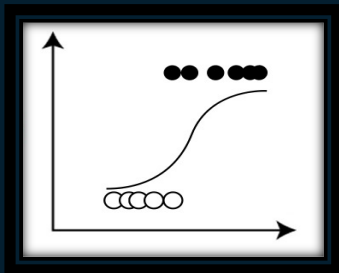
Preprocessat

- ✓ Dades categòriques.
- ✓ Eliminació de valors nulls.
- ✓ Normalització.
- ✓ Variable objectiu.
- ✓ Variables predictores.

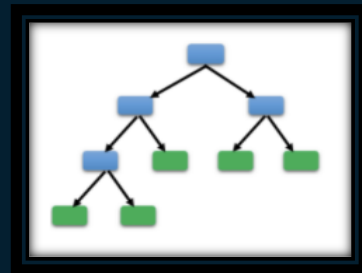


Mètodes utilitzats

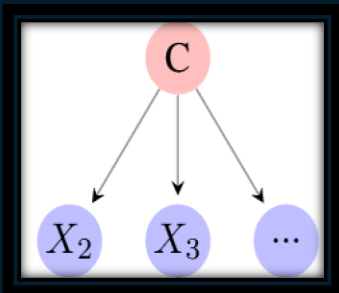
Regressió Logística



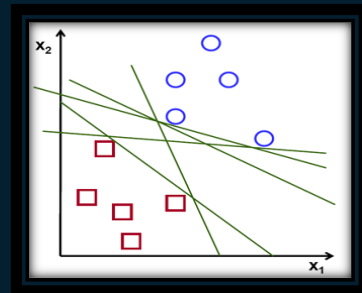
Random Forest



Naive Bayes



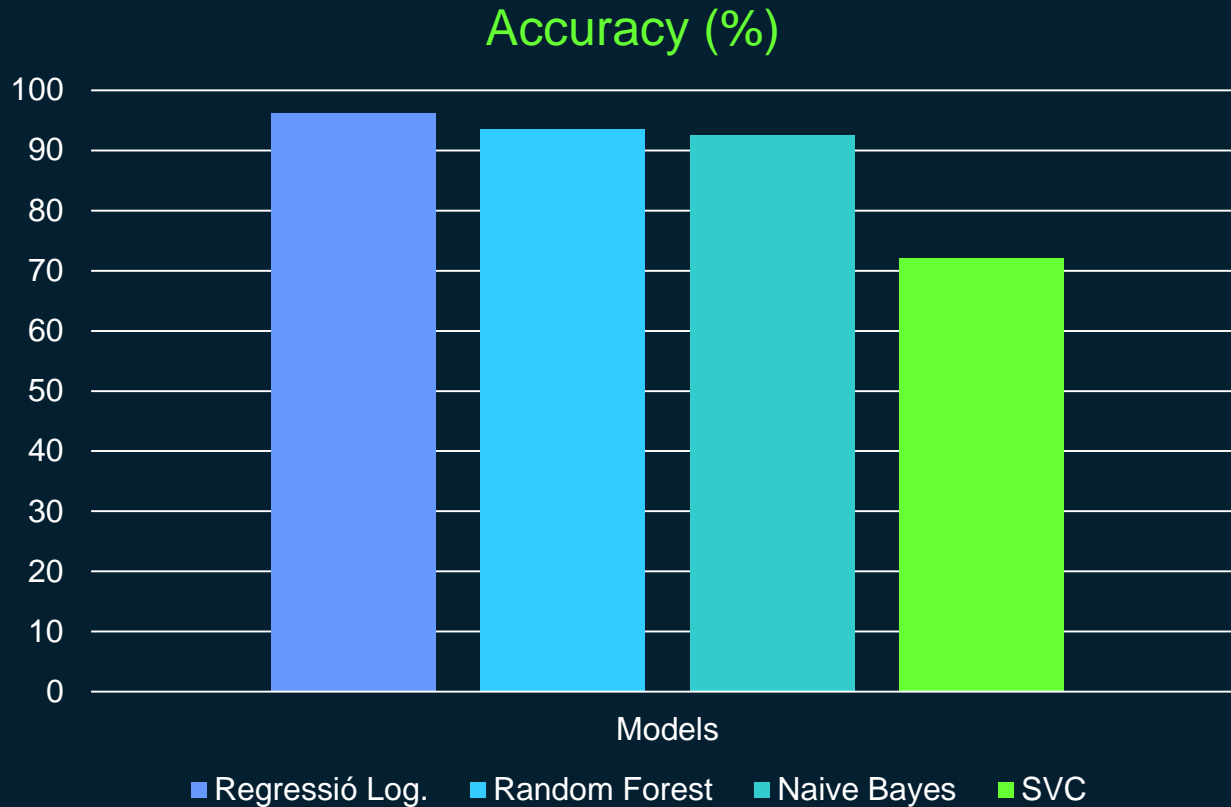
SVC



Resultats

	Accuracy	F1-score	RMSE
Regressió logística	96.26%	93.22%	0.037
Random Forest	93.55%	88.00%	0.064
Naive Bayes	92.59%	86.64%	0.074
SVM	72.03%	01.80%	0.279

Resultats



Conclusions

- Anàlisi de models:
 - ✓ Regressió logística
 - ✓ Random Forest
 - ✓ Naive Bayes
 - ✗ SVC
- Treball a futur (estudi amb nous atributs).

