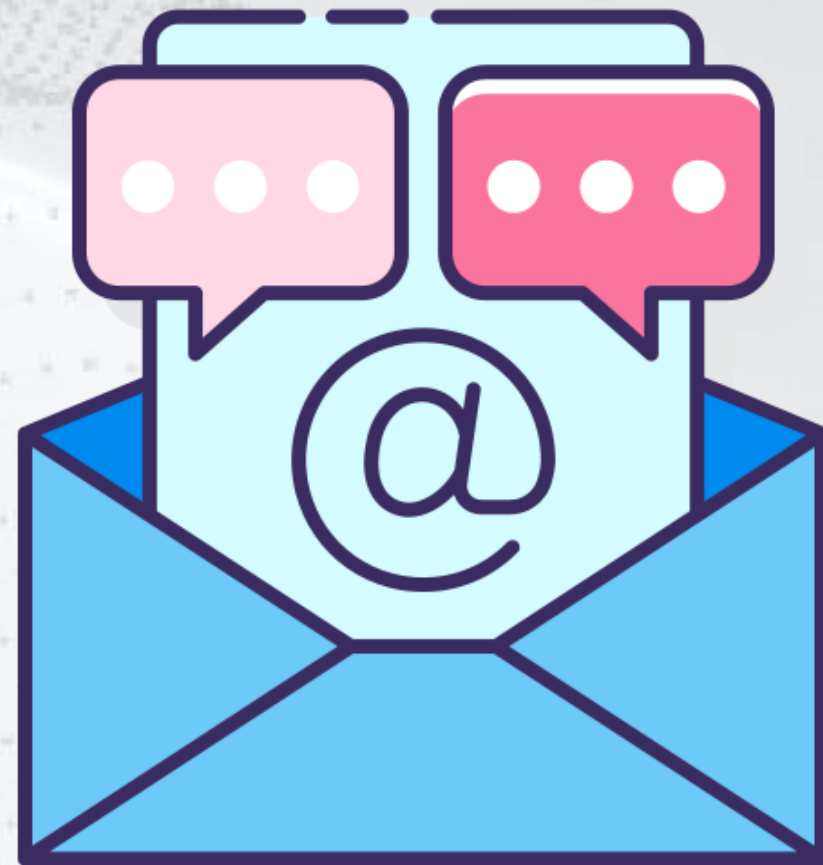


Módulo 5
Clase 11A


Introducción a la Ciencia de Datos

Tuning Modelo Email Spam



Caso Email Spam

- El objetivo de esta actividad aplicar las técnicas para la optimización del modelo de clasificación.



UCI
Machine Learning Repository
Center for Machine Learning and Intelligent Systems

Spambase Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Classifying Email as Spam or Non-Spam

Deleted Items	
ID #	Subject
100	CarLoanProm... Get the car of your dreams with CarLoanProvider help
101	TotalResponse... How Old Are You Really? - Take the Knowledge Test
102	@ Don't be Lame!... (X)Only way to make it great!!
103	Best of Internet... since 1995 - 10-11-11
104	WardrobeShop... Special To The Ladies Member Offer
105	Accept Credit... Process Credit Cards for Zero Up Front Cost
106	Seems... Your Pharmacy is
107	Quick Cash-A... Get a \$500 Cash Advance
108	Leased Delivery... trouble-free delivery
109	edible land... Office 101 - 101
110	Comp Dept... Get a complimentary Starbucks Gift Card on us
111	Guadalupe N... Pay NO Attention to the Man Behind the Curtain
112	Summit Media... Get ready for Monday OCTOBER 15TH

Data Set Characteristics:	Multivariate	Number of Instances:	4601	Area:	Computer
Attribute Characteristics:	Integer, Real	Number of Attributes:	57	Date Donated	1999-07-01
Associated Tasks:	Classification	Missing Values?	Yes	Number of Web Hits:	514344

<https://archive.ics.uci.edu/ml/datasets/spambase>

Instrucciones

- Construya un notebook ordenado, reproducible y documentado
- Aplique técnicas de balanceo de datos
- Aplique técnicas de optimización de hiperparámetros
- Utilice parámetros de regularización
- Compare los resultados del modelo antes de la optimización y después de la optimización
- Elabore conclusiones



¡Gracias!