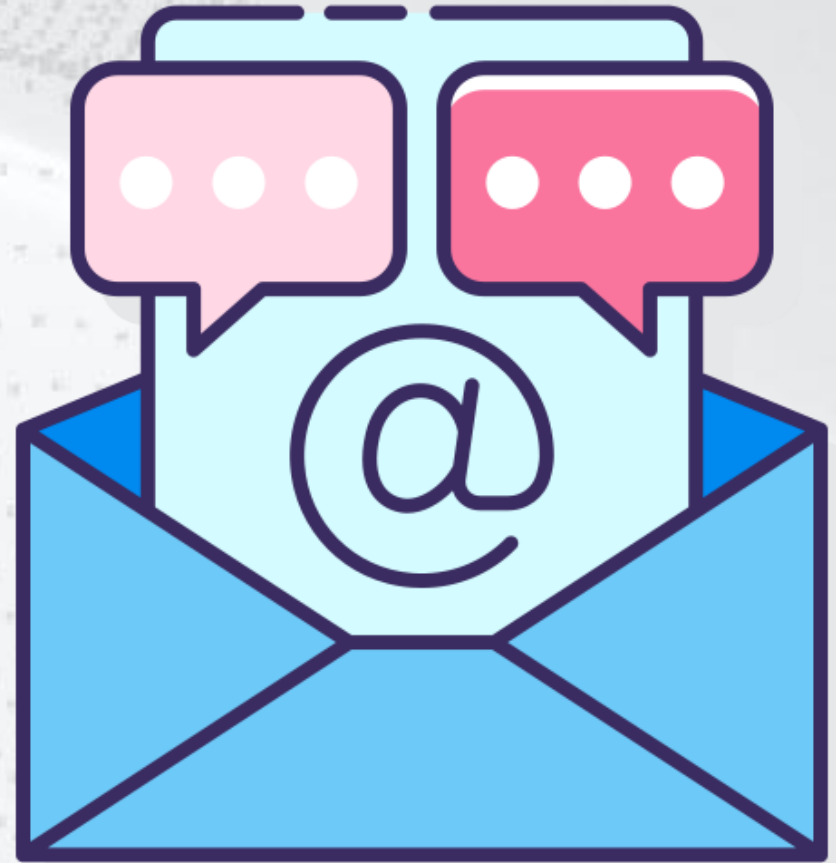


Módulo 5 – Aprendizaje de Máquina Supervisado

# Actividad: Optimización Modelo Email Spam

Especialización en Ciencia de Datos

Actividad:  
Optimización Modelo Email  
Spam



# Caso Email Spam

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



## Spambase Data Set

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

Abstract: Classifying Email as Spam or Non-Spam

| Deleted Items   |  |
|-----------------|--|
| From            | Subject  |
| CarLoanPro...   | Get the car of your dreams with CarLoanPro.com help! |
| SofaResponse... | How Old Are You Really? - Take the RealAge Test!     |
| @ Dorothy L...  | [-] Only way to make it grow!!                       |
| Beryl membe...  | Unit 1 or 2 or 3 or 4 or 5                           |
| Wardlaw, S...   | Special 1st Purchase Member Offer                    |
| Accept Credi... | Process Credit Cards for Zero Up Front Cost          |
| Senex           | Your Pharmacy is...                                  |
| Quick Cash A... | Get a \$1000 Cash Advance                            |
| Levard Denry... | Brandfast employment                                 |
| eddye ind       | Office OP - \$60                                     |
| Comp Dept       | Get a complimentary Starbucks gift Card on us        |
| Guadalupe N...  | Please Attention to the Plan Behind the Curtain      |
| Sunset Meds...  | Get ready for Monday OCTOBER 18th                    |

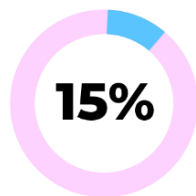
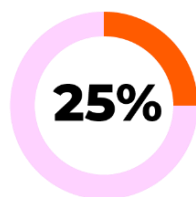
|                            |                |                       |      |                     |            |
|----------------------------|----------------|-----------------------|------|---------------------|------------|
| Data Set Characteristics:  | Multivariate   | Number of Instances:  | 4601 | 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.



¡Éxito!