

Módulo 6 – Aprendizaje de Máquina No Supervisado

Actividad: Caso Tiroides

Especialización en Ciencia de Datos

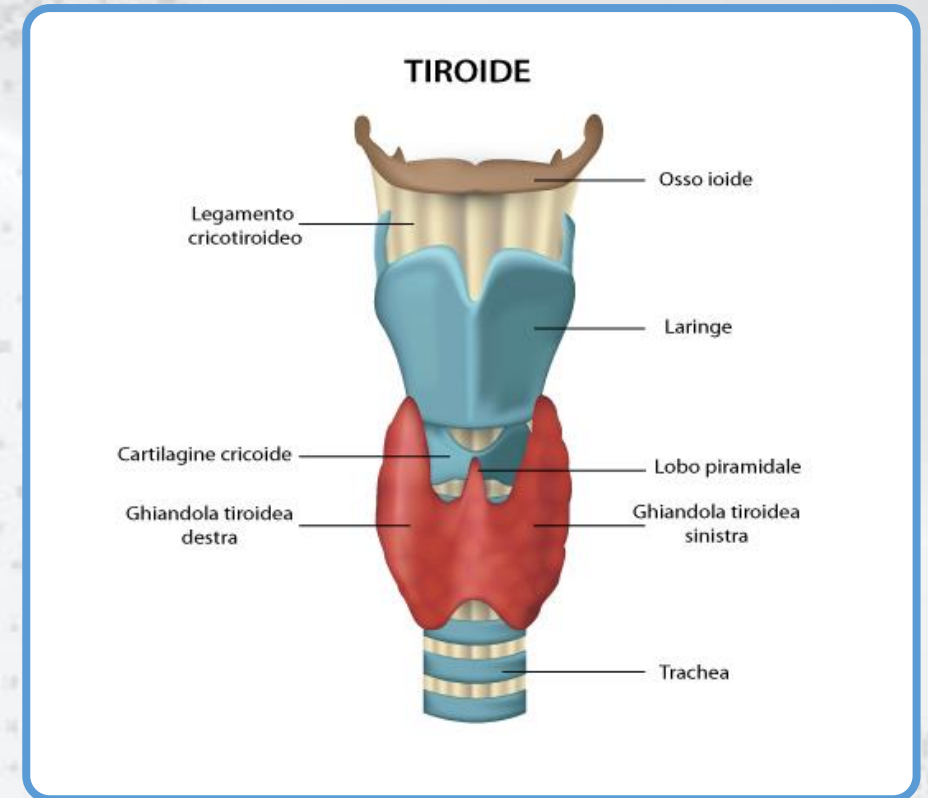
Actividad:Caso Estudio Clínico Enfermedades de la Tiroides



Caso Estudio Clínico Enfermedades de la Tiroides

La incidencia de enfermedades de la tiroides ha aumentado en los últimos tiempos. La glándula tiroides tiene una de las funciones más importantes en la regulación del metabolismo. Las irregularidades en la glándula tiroides pueden dar lugar a diferentes anomalías; dos de los más comunes son el hipertiroidismo y el hipotiroidismo. Un gran número de personas son diagnosticadas anualmente con enfermedades de la tiroides como éstas.

Una predicción proactiva de la enfermedad de la tiroides es esencial para tratar adecuadamente al paciente en el momento adecuado y salvar vidas humanas y gastos médicos. Debido a los avances tecnológicos en el procesamiento y la computación de datos, se aplican técnicas de aprendizaje automático y aprendizaje profundo para predecir el diagnóstico de tiroides en las primeras etapas y clasificar los tipos de enfermedades de la tiroides, hipotiroidismo, hipertiroidismo, etc.



Caso Estudio Clínico Enfermedades de la Tiroides

El set de datos contiene 7.200 mediciones de 21 features ,tanto categóricas como numéricas, que fueron obtenidos en un estudio clínico. Los datos se encuentran normalizados. Adicionalmente, el set de datos cuenta con una columna que clasifica si el paciente bajo estudio presenta o no una enfermedad de la tiroides.

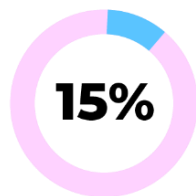
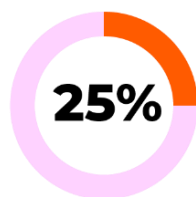
	class	Dim_0	Dim_1=0	Dim_2=0	Dim_3=0	Dim_4=0	Dim_5=0	Dim_6=0	Dim_7=0	Dim_8=0	Dim_9=0	Dim_10=0	Dim_11=0	Dim_12=0	Dim_13=0	Dim_14=0	Dim_15=0	Dim_16	Dim_17	Dim_18	Dim_19	Dim_20
1	0	0.75	1	0	1	1	1	1	1	0	1	1	1	1	1	1	1	0.001132	0.08078	0.197324	0.300926	0.225
2	0	0.239583	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.000472	0.164345	0.235786	0.537037	0.165625
3	0	0.479167	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.003585	0.130919	0.167224	0.527778	0.11875
4	0	0.65625	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.001698	0.091922	0.125418	0.337963	0.129688
5	0	0.229167	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.000472	0.142061	0.229097	0.337963	0.235938
6	0	0.708333	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.000472	0.086351	0.140468	0.24537	0.189063
7	0	0.875	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.000472	0.125348	0.210702	0.402778	0.185937
8	0	0.489583	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.003925	0.108635	0.140468	0.282407	0.16875
9	0	0.6875	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0.002453	0.130919	0.14214	0.425926	0.121875
10	0	0.78125	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.000189	0.158774	0.204013	0.513889	0.148438
11	1	0.635417	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0.020755	0.041783	0.118729	0.263889	0.15
12	0	0.177083	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0.000189	0.125348	0.160535	0.314815	0.176563
13	0	0.604167	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.001509	0.125348	0.153846	0.37963	0.144922
14	0	0.5	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0.001132	0.125348	0.185619	0.393519	0.170313
15	0	0.541667	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0.00434	0.108635	0.102007	0.361111	0.1
16	0	0.395833	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0.000189	0.097493	0.147157	0.25	0.19375
17	0	0.395833	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0.001132	0.108635	0.187291	0.384259	0.175
18	0	0.666667	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0.003019	0.097493	0.12709	0.347222	0.129688
19	0	0.65625	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0.060377	0.075209	0.138796	0.458333	0.107812

Caso Estudio Clínico Enfermedades a la Tiroides

El desafío consiste en plantear ambos enfoques para resolver el problema. Un primer enfoque de aprendizaje supervisado (clasificación) y otro de aprendizaje no supervisado (detección de anomalías) para la elaboración del modelo predictivo. Se pide:

- Elaborar un **modelo supervisado de clasificación** para detectar a los pacientes con enfermedad.
- Elaborar un **modelo no supervisado de detección de anomalías** para detectar a los pacientes con enfermedad.
- Utilice Orange Mining.





¡Éxito!