



## INGA7\_02.pdf

Integrantes: Stiven Alomaliza, Anthony Galindo, Bryan Llamuca  
Universidad Técnica de Ambato  
Tecnologías de la Información

### Introducción

El presente informe tiene como objetivo realizar un análisis exhaustivo de datos relacionados con síntomas médicos y diagnósticos correspondientes, así como la construcción de un árbol de decisión basado en estos datos. El análisis y la construcción del árbol de decisión se realizaron siguiendo una metodología estructurada para identificar patrones, establecer reglas de diagnóstico y verificar la efectividad del modelo resultante.

### Análisis de los datos para identificar patrones y relaciones

Vamos a descargar el archivo diagnóstico.csv Impórtelo a Excel Analice los registros de cada uno de los pacientes Establezca las posibles relaciones entre el diagnóstico y los síntomas.

	A	B	C	D	E	F	G
1	Patient ID	Dolor de garganta	Fiebre	Glándulas inflamadas	Congestión	Dolor de cabeza	Diagnóstico
2	1	Si	Si	Si	Si	Si	Estreptococo
3	2	No	No	No	Si	Si	Alergia
4	3	Si	Si	No	Si	No	Resfriado
5	4	Si	No	Si	No	No	Estreptococo
6	5	No	Si	No	Si	No	Resfriado
7	6	No	No	No	Si	No	Alergia
8	7	No	No	Si	No	No	Estreptococo
9	8	Si	No	No	Si	Si	Alergia
10	9	No	Si	No	Si	Si	Resfriado
11	10	Si	Si	No	Si	Si	Resfriado
12							

Se examinaron los datos proporcionados en el archivo de Excel, que consisten en información sobre la presencia o ausencia de síntomas como dolor de garganta, fiebre, glándulas inflamadas, congestión y dolor de cabeza, junto con el diagnóstico correspondiente.

Se observó la frecuencia de cada síntoma y su relación con los diferentes diagnósticos.

Se utilizaron herramientas de visualización de datos para identificar patrones más claramente, lo que permitió una comprensión más profunda de las relaciones entre los síntomas y los diagnósticos.



### Encontrar reglas que justifiquen el diagnóstico basado en el análisis de los datos

Se identificaron reglas lógicas que relacionan los síntomas observados con los diagnósticos proporcionados.

Por ejemplo, se estableció que, si un paciente presenta dolor de garganta, fiebre y glándulas inflamadas, es probable que tenga estreptococo.

Tenemos que si el paciente presenta congestión y no presenta otro síntoma es más probable que sea Alergia.

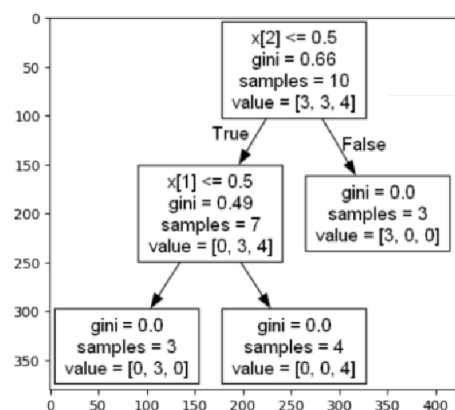
Si tenemos fiebre y congestión es probable que tenga resfriado

Estas reglas se basaron en el análisis de los datos disponibles y en su validez clínica.

Para utilizar el análisis de datos dentro de Excel vamos a tener sus datos numéricos ya sea que para Yes = 1 y No = 0 vamos a ver cambiar para Estreptococo = 0, Alergia = 1, Resfriado = 2.

patient_id	dolor_garga	fiebre	glans_inflama	congesti	dolo_cabe	diagnos
1	1	1	1	1	1	0
2	0	0	0	1	1	1
3	1	1	0	1	0	2
4	1	0	1	0	0	0
5	0	1	0	1	0	2
6	0	0	0	1	0	1
7	0	0	1	0	0	0
8	1	0	0	1	1	1
9	0	1	0	1	1	2
10	1	1	0	1	1	2

### Construir un árbol de decisión utilizando las reglas encontradas.





Se utilizó un enfoque de árbol de decisión para organizar las reglas identificadas en un modelo comprensible y fácil de interpretar. Los síntomas se organizaron en nodos del árbol, y se definieron las condiciones que conducen a cada diagnóstico. Se aplicaron técnicas de poda para simplificar el árbol y mejorar su generalización.

### **Conclusiones**

El análisis de datos y la construcción de un árbol de decisión proporcionaron un marco sólido para la clasificación de diagnósticos médicos basados en síntomas observados. Las reglas identificadas y el árbol de decisión resultante demostraron ser efectivos para predecir diagnósticos en nuevos conjuntos de datos. Sin embargo, se reconoce la importancia de continuar refinando y validando el modelo para garantizar su precisión y generalización en diversas situaciones clínicas.