

**UNIVERSIDAD MAYOR DE SAN ANDRÉS**  
**FACULTAD DE CIENCIAS PURAS Y NATURALES**  
**CARRERA DE INFORMÁTICA**

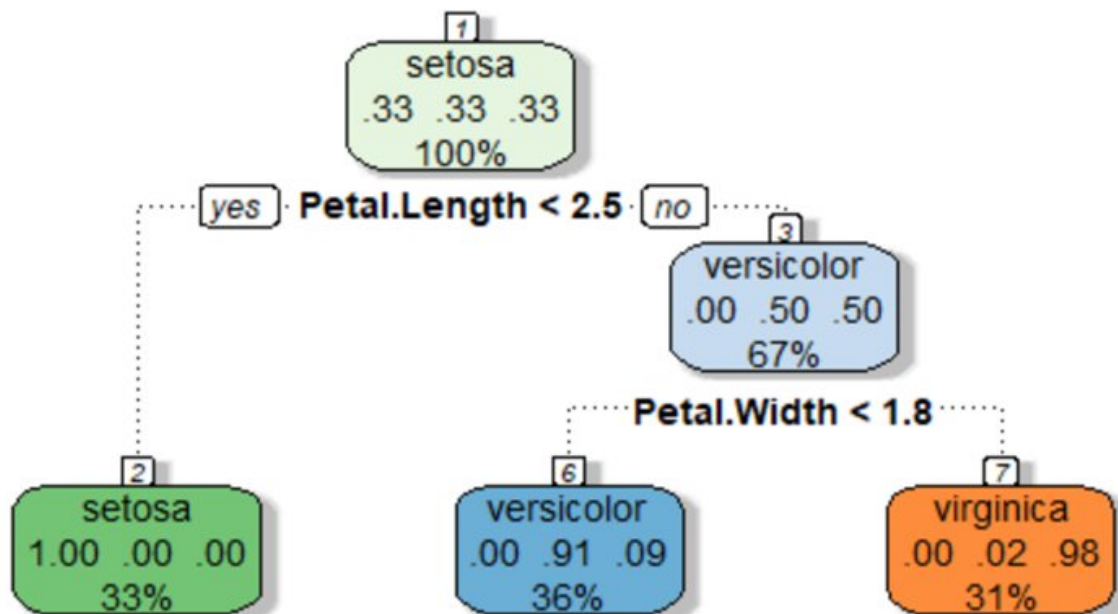


**CLASIFICADOR: ARBOL DE DESICION**

<b>Dataset</b>	Diabetes Health Indicators
<b>Autor/es</b>	<b>Nombres y Apellidos</b>
	Gary Omar Nova Mamani
<b>Fecha</b>	16/06/2023
<b>Carrera</b>	Informática
<b>Asignatura</b>	INF - 354
<b>Docente</b>	Lic. Moises Silva



# Introduccion



El uso de los arboles de desicion para el dataset “Diabetes Health Indicators” puede proporcionar una serie de beneficios cuando nos encontramos analizando la enfermedad y que un arbol de desicion puede manejar mejor las características y clasificar segun vea conveniente.

Un arbol de desicion ofrece una mejor adaptabilidad. aplicabilidad, interpretabilidad y una capacidad de manejos de datos complejos asi como para identificar características relevantes.

## Algunas razones son:

- **Aplicabilidad en el campos de la diabetes:** Segun articulos sobre investigaciones y predicciones de modelos usados para la clasificacion de la diabetes podemos encontrar que los **Arboles de desicion** son muy usados en campos medicos ya dentro de estos podemos encontrar una gran capacidad de abordar problemas de clasificacion y prediccion ya que estos modelos pueden usados para diagnosticos tempranos para el pronosticos de la enfermedad asi como la identificacion de factores de riesgo.
- **Interpretabilidad:** Los árboles de decisión son claramente interpretables, esto significa que sus resultados pueden ser fácilmente comprendidos y explicados. Cada nodo y rama del árbol representa una decisión basada en una característica específica del dataset. Esto hace que cada persona que lea el dataset, podra comprender cada paso que se dio dentro del arbol de desicion y cómo se llega a una determinada clasificación o predicción, los factores de riesgo y las recomendaciones de tratamiento.
- **Manejo de datos complejos:** El dataset de diabetes a menudo contiene una variedad de variables clínicas y biomédicas, como edad, índice de masa corporal, niveles de



glucosa en sangre, presión arterial, entre otros. Los árboles de decisión son capaces de manejar datos complejos y heterogéneos, incluyendo tanto variables numéricas (1,0) como categóricas. Esto significa que no se requiere una preparación extensa de los datos antes de aplicar el modelo, lo que ahorra tiempo y esfuerzo en la etapa de preprocesamiento.

- **Identificación de características relevantes:** Los árboles de decisión pueden identificar las características más relevantes o predictivas en el dataset de diabetes. Al analizar la estructura del árbol, es posible determinar qué variables tienen un mayor impacto en la toma de decisiones. Esta información puede ser crucial para la investigación médica y la comprensión de los factores de riesgo asociados con la diabetes.
- **Interpretación de resultados:** Los árboles de decisión permiten la interpretación de resultados a través de diferentes medidas, como la importancia de las características y la precisión del modelo. Estas medidas pueden ayudar a los profesionales de la salud o personas a fines a evaluar la confiabilidad y la utilidad del modelo en el contexto de la diabetes.

Como podemos ver el arbol de desicion puede ayudarnos mucho en el manejo de datos mucho mas complejos, y hablando del dataset de la diabetes podemos decir que la clasificacion con este modelo es mucho mas eficiente, completa e interpretable para aquellos que quieran entender y saber como se llega a las predicciones. Segun Nongyao Nai-arun en su estudio de los tipos de calsificadores para un dataset de diabetes que cuenta con los siguientes campos: BMI, edad, peso, waist\_cm, Bph, bpl, si algun familiar sufre de diabetes, si algun familiar tiene algun historial de hipertension, alcohol, fuma, sexo. El arbol de desicion obtuvo una mayor tasa de rendimiento conforme a los datos con una presicion de 85.090% frente a los otros modelos de clasificacion. (<https://doi.org/10.1016/j.procs.2015.10.014>)

Segun el articulo de “Comparison of Classifiers for the Risk of Diabetes Prediction” cuando contamos con datos que tienen un enfoque supervisado, nosotros podemos usar metodos de clasificacion supervisada, donde las variables son dependientes y donde queremos es hayar el objetivo correcto con las nuevas clases.

(<https://pistaseducativas.celaya.tecnm.mx/index.php/pistas/article/view/1805/1471>)

Segun Raul lopez Fernandez respecto al porcentaje de clasificación, los modelos se comportaron de forma similar, aunque el árbol de decisión con CHAID es ligeramente superior con un 80 % de clasificación correcta. Respecto al VPP los modelos mostraron valores adecuados, superiores en el árbol de decisión con CHAID, que fue de 85 %. Los resultados de la sensibilidad, tanto para el modelo de regresión logística, como para el árbol de decisión, se comportaron de forma similar, ya que ambos clasificaron correctamente a los pacientes que tenían padecimiento de pie diabético con valores de 89,3 % y 89,1 %, respectivamente. (ISSN 1727-897X)



## Referencias

- Ollé, J. (2020b, February 20). *¿Qué es un Árbol de Decisión y cómo interpretarlo?* Conceptos Claros; Jordi Ollé. <https://conceptosclaros.com/que-es-arbol-decision/>
- López Fernández, R., Yanes Seijo, R., Suárez Surí, P., Avello Martínez, R., Gutiérrez Escobar, M., & Alvarado Flores, R. (2016). Modelo estadístico predictivo para el padecimiento de pie diabético en pacientes con diabetes mellitus tipo II. *Medisur*, 14(1), 42–52. [http://scielo.sld.cu/scielo.php?pid=S1727-897X2016000100009&script=sci\\_arttext&tlng=en](http://scielo.sld.cu/scielo.php?pid=S1727-897X2016000100009&script=sci_arttext&tlng=en)
- May, O. A. C., Koo, J. J. P., Kinani, J. M. V., & Encalada, M. A. Z. (2018). CONSTRUCCIÓN DE UN MODELO DE PREDICCIÓN PARA APOYO AL DIAGNÓSTICO DE DIABETES (CONSTRUCTION OF A PREDICTION MODEL TO SUPPORT THE DIABETES DIAGNOSIS). *Pistas Educativas*, 40(130). <https://pistaseducativas.celaya.tecnm.mx/index.php/pistas/article/view/1805/1471>