



TRABAJO FINAL

MODELO ANALÍTICO



Equipo 2



**Aranda
Andrea**



**Bravo
Maria**



**García
Isaac**



**González
Lucero**



**Morales
Samuel**



**Romero
Esteban**



**Ruiz
Angel**

Explicación del Data Frame

	id	gender	age	hypertension	heart_disease	ever_married	work_type	avg_glucose_level	bmi	smoking_status	stroke
0	9046	1	67.0	0	1	1	1	228.69	36.6	0	1
1	51676	0	61.0	0	0	1	2	202.21	NaN	1	1
2	31112	1	80.0	0	1	1	1	105.92	32.5	1	1
3	60182	0	49.0	0	0	1	1	171.23	34.4	2	1
4	1665	0	79.0	1	0	1	2	174.12	24.0	1	1
...
5105	18234	0	80.0	1	0	1	1	83.75	NaN	1	0
5106	44873	0	81.0	0	0	1	2	125.20	40.0	1	0
5107	19723	0	35.0	0	0	1	2	82.99	30.6	1	0
5108	37544	1	51.0	0	0	1	1	166.29	25.6	0	0
5109	44679	0	44.0	0	0	1	3	85.28	26.2	3	0

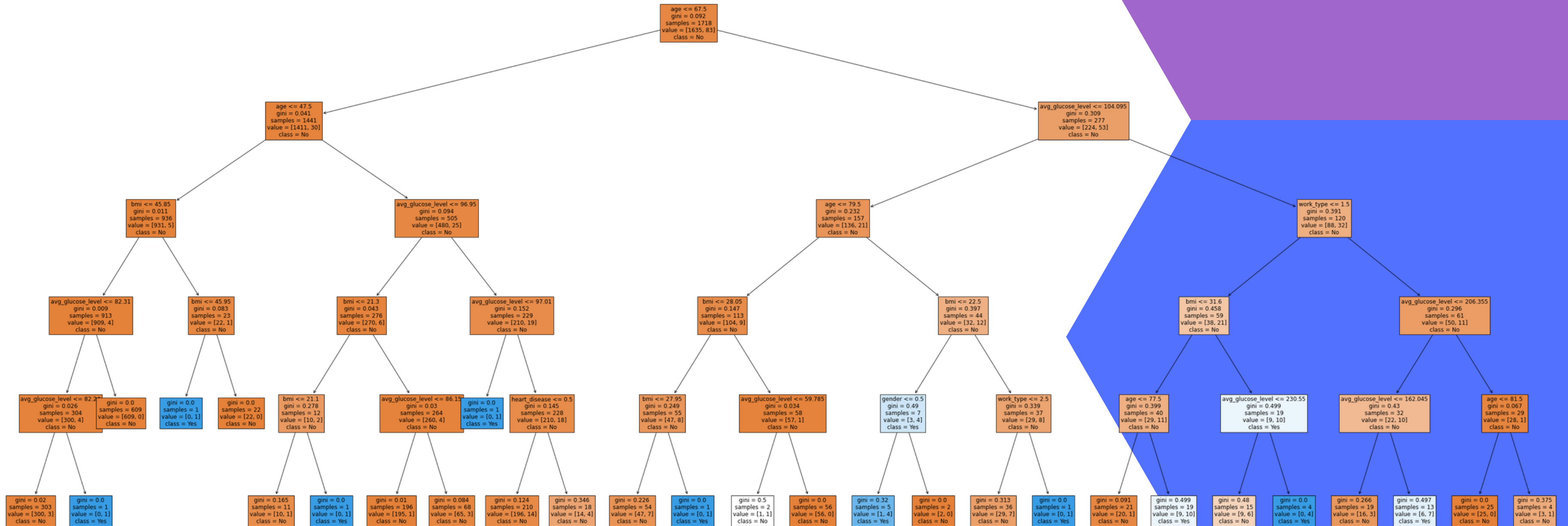
Private:1, Self-employed:2, Govt_job:3, children:4, Never_worked:5

Yes:1, No:0

Female:0, Male:1, Other:2

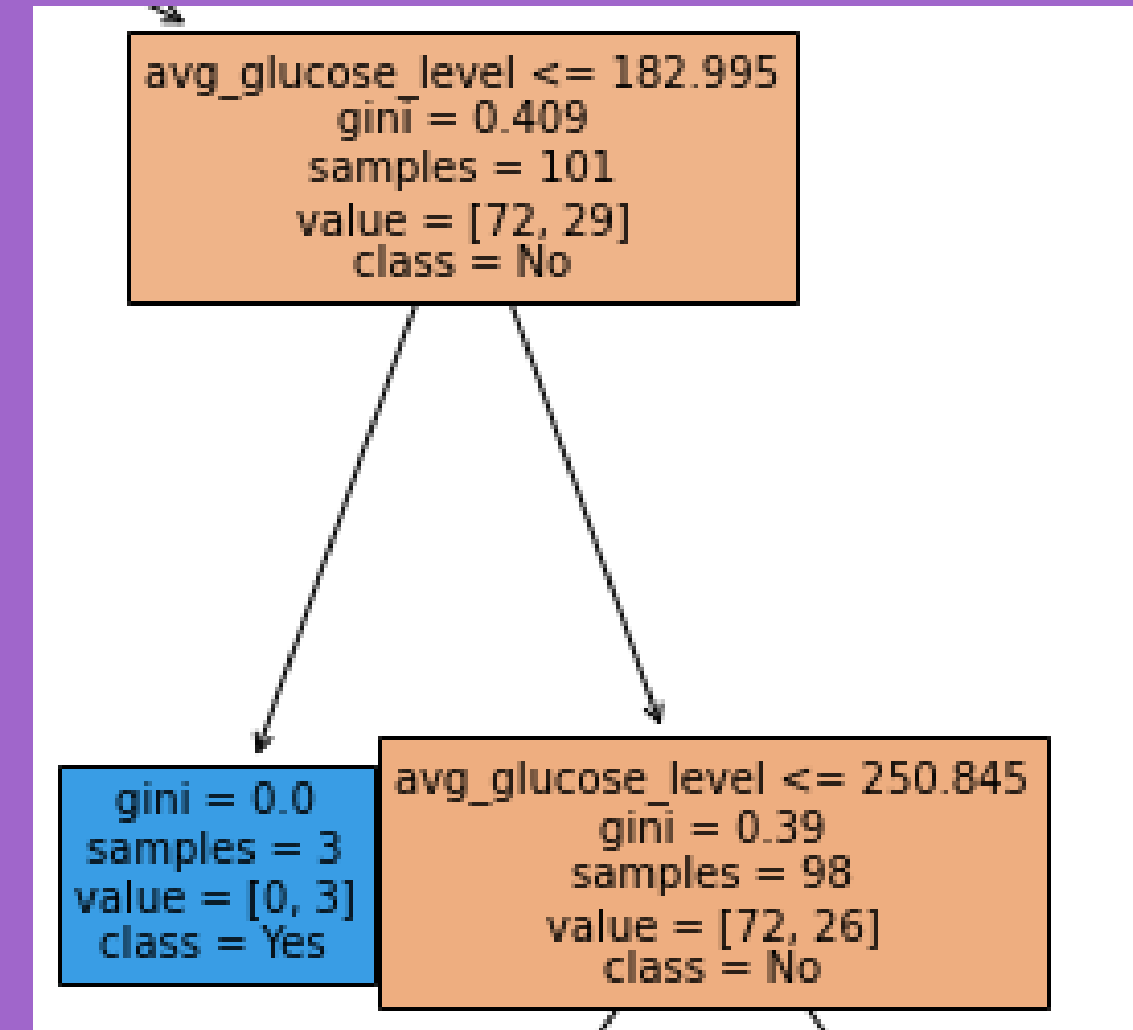
formerly smoked:1, never smoked:0, smokes:2, Unknown:3

Modelo de árbol de decisión



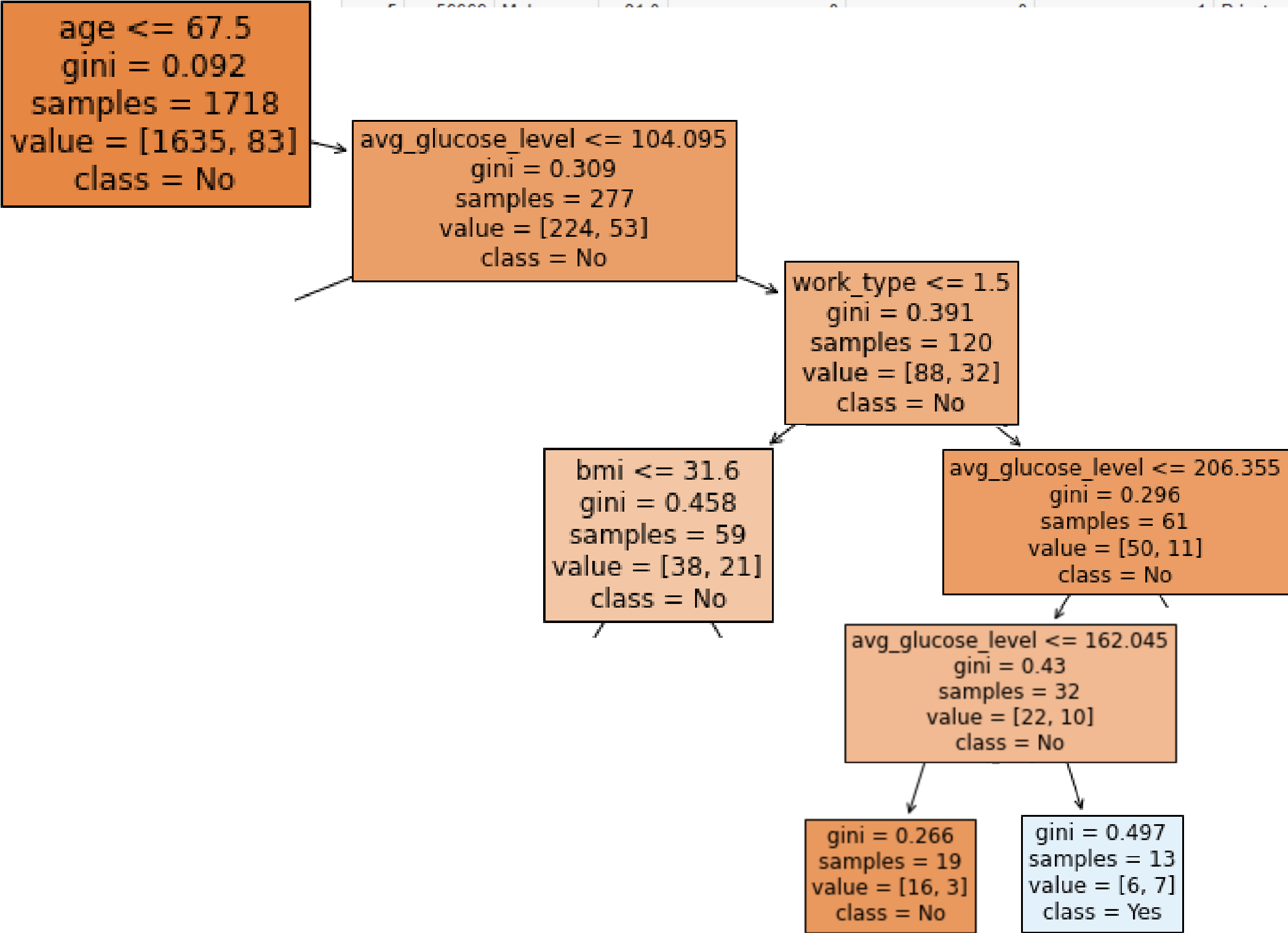
Interpretación del árbol de decisión

- condición: es el primer renglón, a la izquierda es verdadero y a la derecha falso.
- gini: es una medida de impureza (0.0 totalmente puro).
- sample: es el número de muestras que satisfacen la condición.
- value: las muestras de cada clase que llegan a ese nodo.
- class: la clase que es le asigna a este nodo



Ejemplo 1

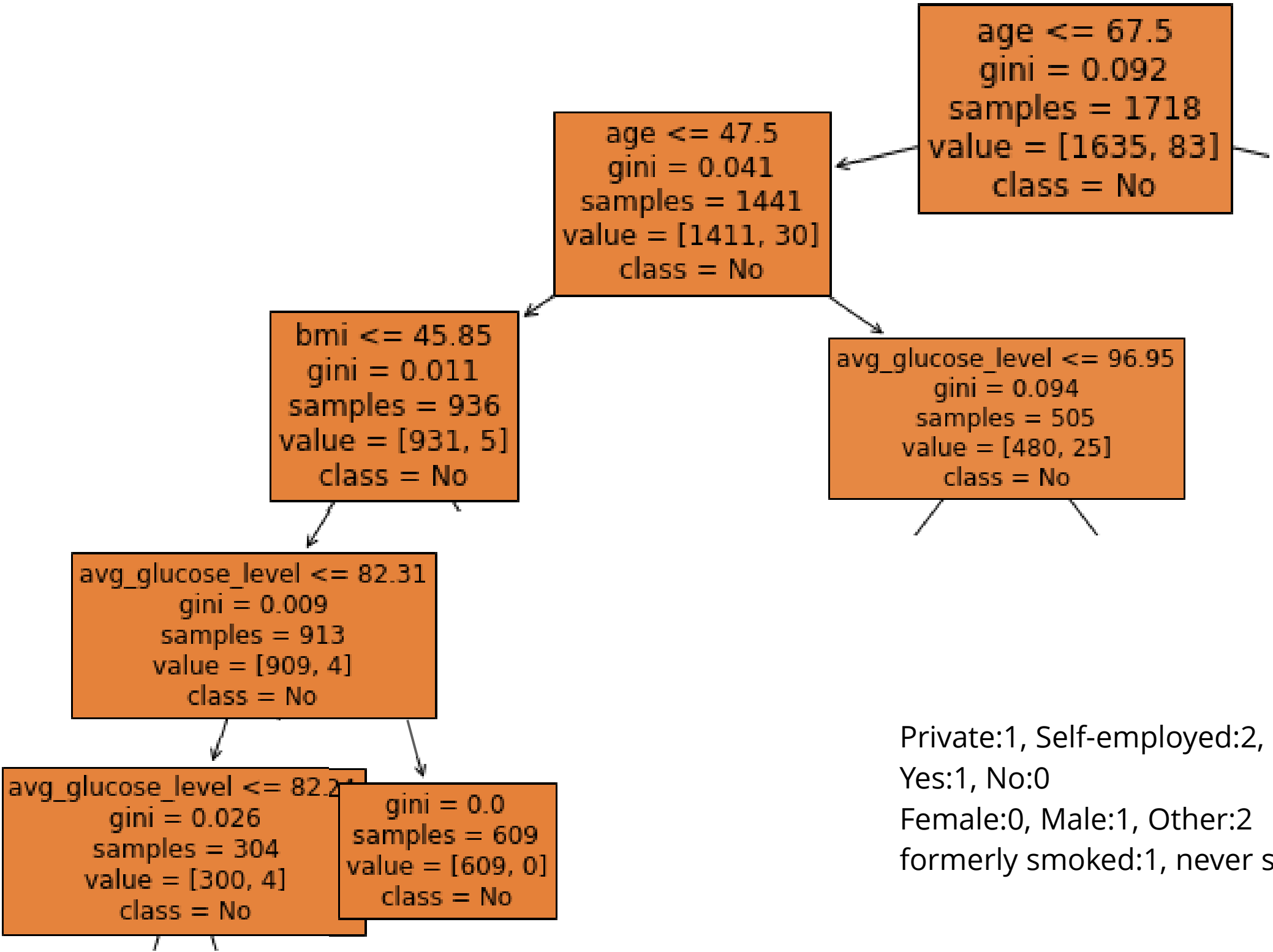
index	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
4	1665	Female	79.0	1	0	1	Self-employed	Rural	174.12	24.0	never smoked	1
5	56888	Male	61.0	0	0	1	Self-employed	Urban	186.34	28.0	formerly smoked	1



Private:1, Self-employed:2, Govt_job:3, children:4, Never_worked:5
Yes:1, No:0
Female:0, Male:1, Other:2
formerly smoked:1, never smoked:0, smokes:2, Unknown:3

Ejemplo 2

index	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
2489	29229	Male	32.0	0	0	1	Private	Urban	92.08	28.4	smokes	0



Private:1, Self-employed:2, Govt_job:3, children:4, Never_worked:5
Yes:1, No:0
Female:0, Male:1, Other:2
formerly smoked:1, never smoked:0, smokes:2, Unknown:3

Recomendaciones del diagrama de árbol

01 Ventajas

Son fáciles de comprender y traducir, pueden trabajar con conjuntos de datos tanto enteros como decimales, pueden trabajar con datos multidimensionales y no requieren establecer parámetros.

02 Desventajas

Los atributos de salida deben ser categorías, no se permiten múltiples atributos de salida, los árboles contruidos a partir de datos numéricos pueden resultar muy complejos

03 ¿Por qué lo usamos nosotros?

Es muy útil para el diagnóstico de enfermedades, de fallos de sistemas, la clasificación de clientes con el fin de aplicar estrategias de marketing o de detectar si hay riesgo en conceder un préstamo.

Conclusiones de los resultados obtenidos

De acuerdo con los análisis realizados a través del modelo de árbol de decisión que generamos se puede predecir si alguna persona de sexo femenino o masculino es propensa a sufrir un derrame o por otro lado ya ha sufrido dicho ataque, esto partiendo de una serie de características como lo son, su historial medico y algunas otras de la forma de vida que lleva dicha persona, a través de la edad, si tiene o no hipertensión, si es fumador o no, los niveles de glucosa y demás datos existentes en el dataframe podemos predecir un derrame, lo cual nos sirve como método de prevención para los pacientes de un hospital por ejemplo, señalándole que modificando algunas conductas de su vida cotidiana puede tener una mejor calidad de vida o darle seguimiento a dicho padecimiento antes de que se agrave.