

Proyecto Minería de datos

Gerardo Moreno	A01734876
Cristina López Ontiveros	A01424566
Sofía Cañas	A01173828
Diego Flores	A00831191

Base de Datos: Diabetes

768 datos

01 | **Pregnancies**
Número de embarazos

02 | **Glucose**
Concentración de Glucosa

03 | **BloodPressure**
Presión arterial mm Hg

04 | **SkinThickness**
Espesor de tricep mm

05 | **Insulin**
Concentración de Insulina

06 | **BMI**
Indice de masa corporal

07 | **DiabetesPedigreeFunction**
Score basado en historial familiar

08 | **Age**
Años

09 | **Outcome**
Resultado diabetes binario (0,1) -> (no, yes)



"La inteligencia artificial no va a sustituir a los médicos. Es un aliado" - José Cerrolaza

Red Neuronal Artificial



Una capa: 30 Neuronas

Activación: Relu

Optimizador: Adam

Batch Size: 32

Epochs = 150

Red Neuronal Artificial Resultados

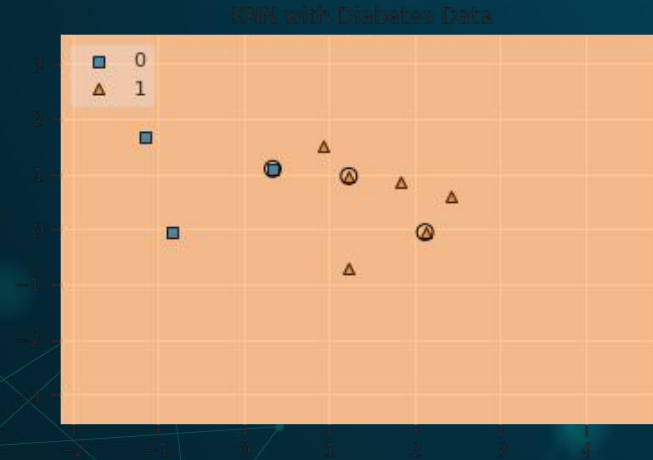
	Precisión	Sensibilidad	Score F1
0	0.83	0.88	0.85
1	0.75	0.66	0.70
Exactitud			0.80

Condición

Predicción

	0	1
0	109	15
1	23	45

K-Nearest Neighbors



```
k = [11]
```

K-Nearest Neighbors

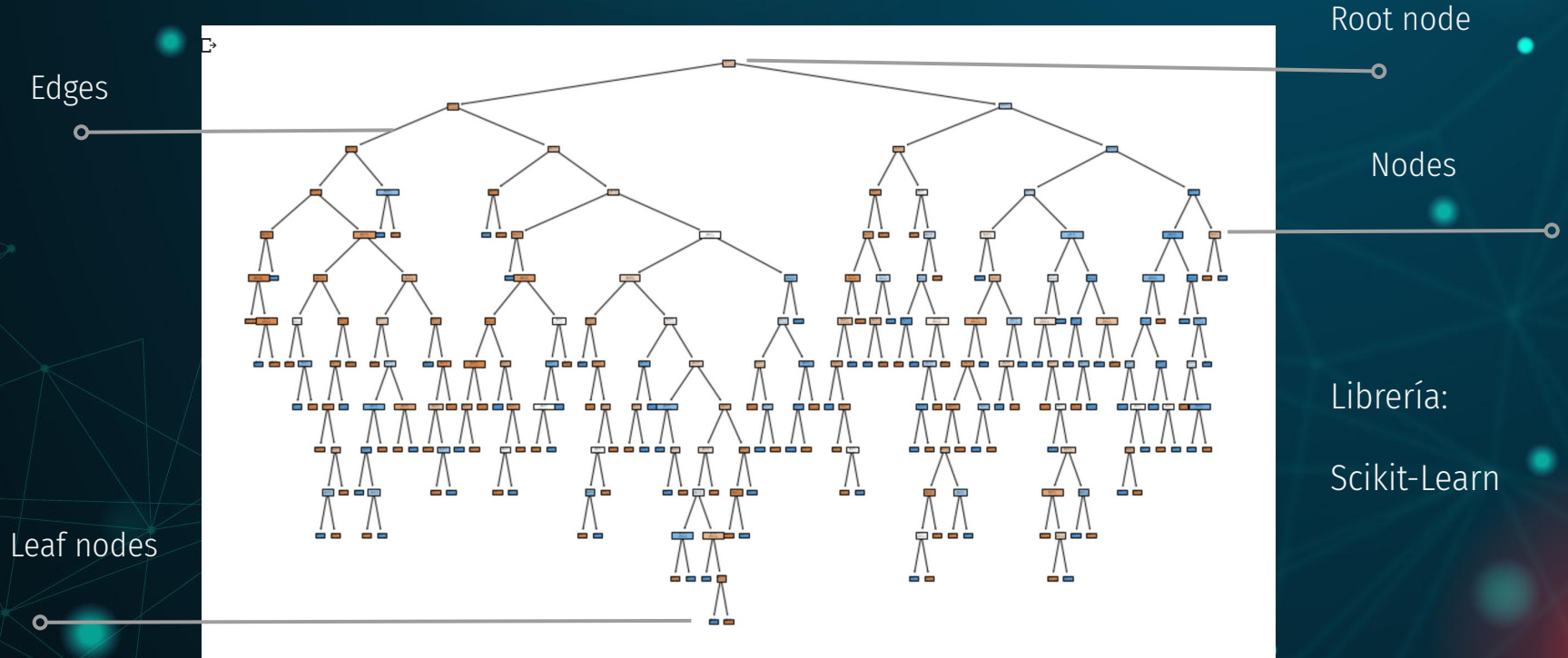
Resultados

	Precisión	Sensibilidad	Score F1
0	0.80	0.85	0.83
1	0.68	0.61	0.64
Exactitud			0.76

Condición

Predicción		
	0	1
0	142	25
1	35	54

Decision Trees



Decision Trees

Resultados

	Precisión	Sensibilidad	Score F1
0	0.80	0.84	0.82
1	0.62	0.56	0.59
Exactitud			0.75

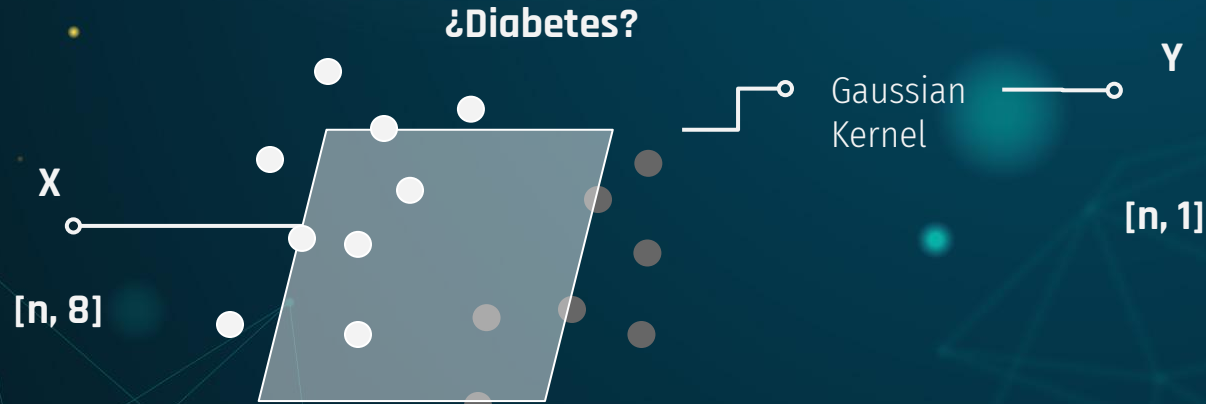
Condición

Predicción		
	0	1
0	74	23
1	22	35

Importancia por cada variable



Support Vector Machine



Hiperplano: 8 dimensiones

Kernel: Gaussiano

Support Vector Machine Resultados

	Precisión	Sensibilidad	Score F1
0	0.77	0.92	0.84
1	0.80	0.55	0.65
Exactitud			0.78

Condición

	Predicción	
	0	1
0	44	14
1	13	16

Comparación Exactitud



Árbol de decisión

0.75



SVC Kernel Gaussiano

0.78



KNN

0.76



Red Neuronal

0.80

Reflexiones

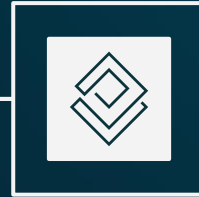
Apreciar la rapidez con la que se puede entrenar un modelo listo para predecir de manera aceptable

Gerardo



Diego

Como los resultados obtenidos en los cuatro modelos son similares será importante fijarse el rapidez y en la memoria utilizada



En un futuro se podría mejorar cada uno de los modelos con un mayor número de datos, así que sería interesante ver el incremento de su exactitud

Cristina

Sofía

Cada modelo nos presentó un acercamiento diferente a la problemática, enseñándonos distintas maneras de analizar los datos y posibles ventajas y desventajas para situaciones específicas



Referencias

Librerías Python

- Pandas
- Numpy
- Scikit-Learn
- Keras



Thanks