

Clasificación de setas según su toxicidad

Gabriel Ordoñez, Tatiana Flórez.
22971- Inteligencia Artificial I - Grupo H1
Escuela de Ingeniería de Sistemas e Informática



CLASIFICACION DE SETAS

Introducción



¿Son **comestibles** o **venenosos**?

Contenidos

01

Dataset

02

Metodología

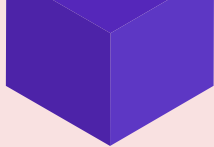
03

Resultados

04

Conclusiones



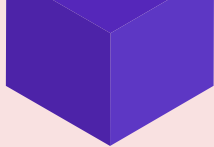


01 Dataset

Mushrooms

Class Edible = e Poisonous = p	Cap-shape Bell = b Conical = c Convex = x Flat = f Knobbed = k Sunken = s	Cap-surface Fibrous = f Grooves = c Scaly = x Smooth = s	Cap-color Brown = n Buff = b Cinnamon = c Gray = g Green = r Pink = p Purple = u red = e White = w Yellow = y	Oddor Almond = a Anise = l Creosote = c Fishy = y Foul = f Musty = m Pungent = p Spicy = s None = n
p	x	s	n	p
e	x	s	y	a
e	b	s	w	l



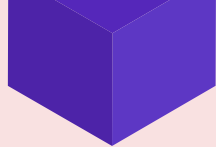


01 Dataset

Mushrooms

 Class Edible = e Poisonous = p	Cap-shape Bell = b Conical = c Convex = x Flat = f Knobbed = k Sunken = s	Cap-surface Fibrous = f Grooves = c Scaly = x Smooth = f	Cap-color Brown = n Buff = b Cinnamon = c Gray = g Green = r Pink = p Purple = u red = e White = w Yellow = y	Oddor Almond = a Anise = l Creosote = c Fishy = y Foul = f Musty = m Pungent = p Spicy = s None = n
p	x	s	n	p
e	x	s	y	a
e	b	s	w	l



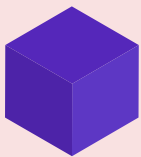


01 Dataset Mushrooms

El dataset original tiene un total de 23 columnas

Class Edible = e Poisonous = p	Cap-shape Bell = b Conical = c Convex = x Flat = f Knobbed = k Sunken = s	Cap-surface Fibrous = f Grooves = c Scaly = x Smooth = f	Cap-color Brown = n Buff = b Cinnamon = c Gray = g Green = r Pink = p Purple = u red = e White = w Yellow = y	Oddor Almond = a Anise = l Creosote = c Fishy = y Foul = f Musty = m Pungent = p Spicy = s None = n
p	x	s	n	p
e	x	s	y	a
e	b	s	w	l

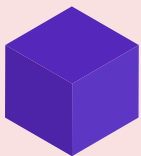




02

Metodología





02

Metodología

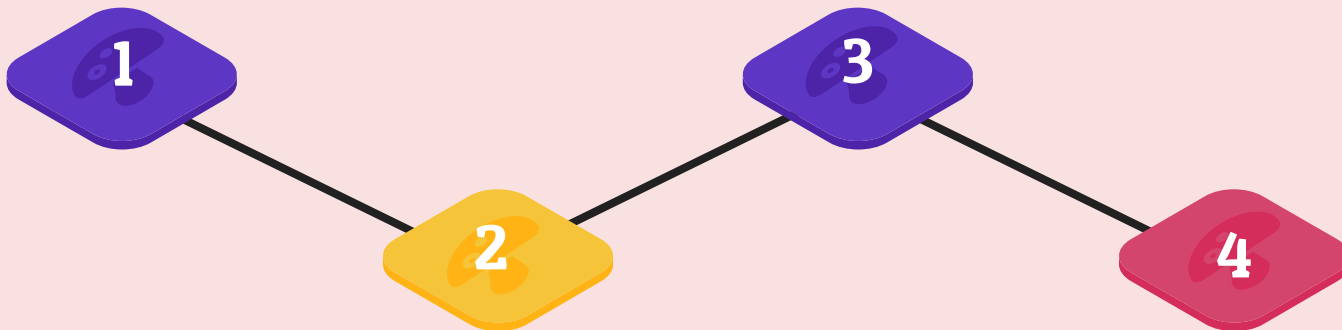


Procesamiento de datos parte 1

(One Hot encoding, Label encoding)

Entrenamiento usando clasificadores

(Naïve Bayes, SVM, DTC, RFC)

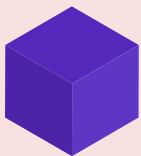


Procesamiento de datos parte 2

(PCA)

Test





02

Metodología

Procesamiento de datos parte 1

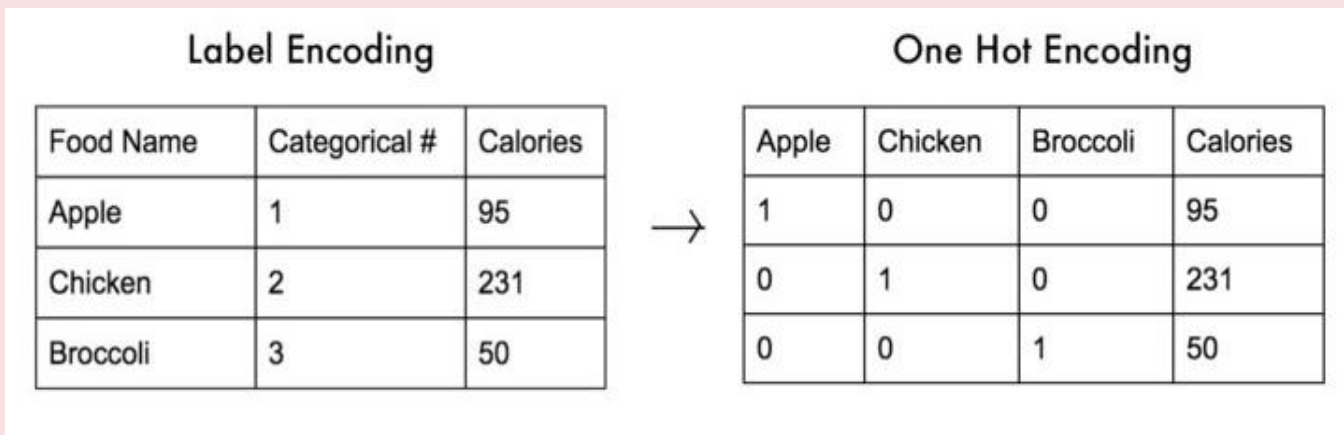
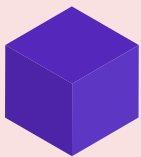


Figura 1. Ejemplo de One hot Encoding.





02

Metodología

Procesamiento de datos parte 2

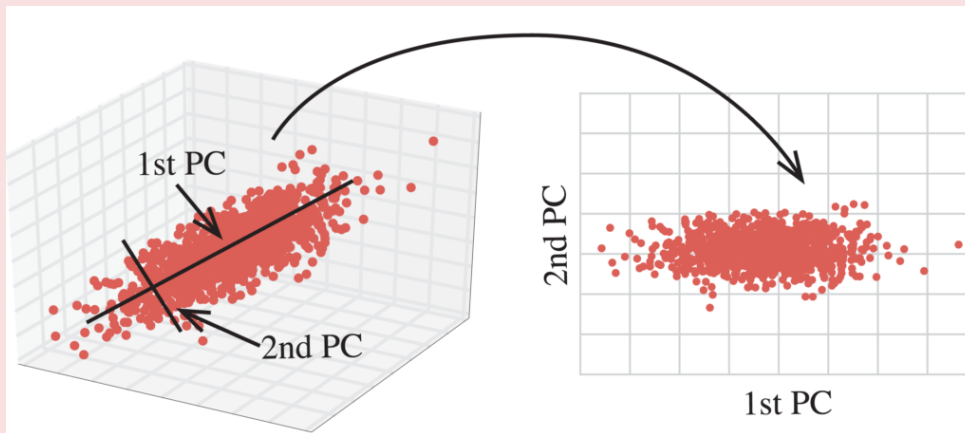
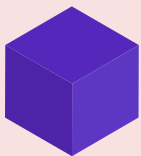


Figura 2. Ejemplo de PCA.





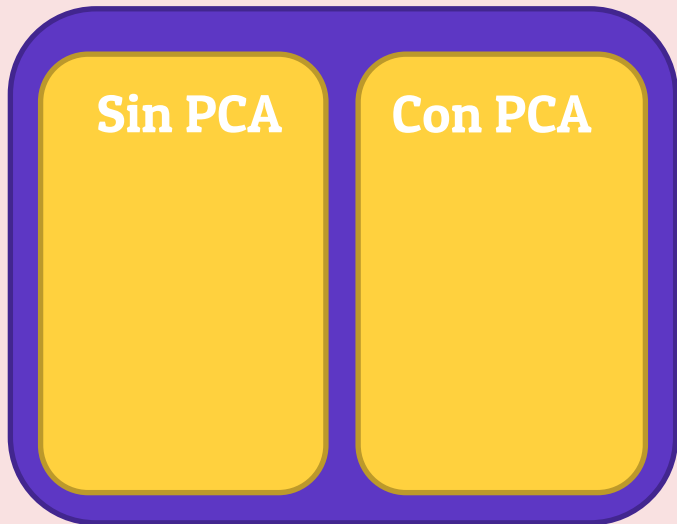
02

Metodología

Entrenamiento usando clasificadores



Clasificadores con LE



Clasificadores con OHE

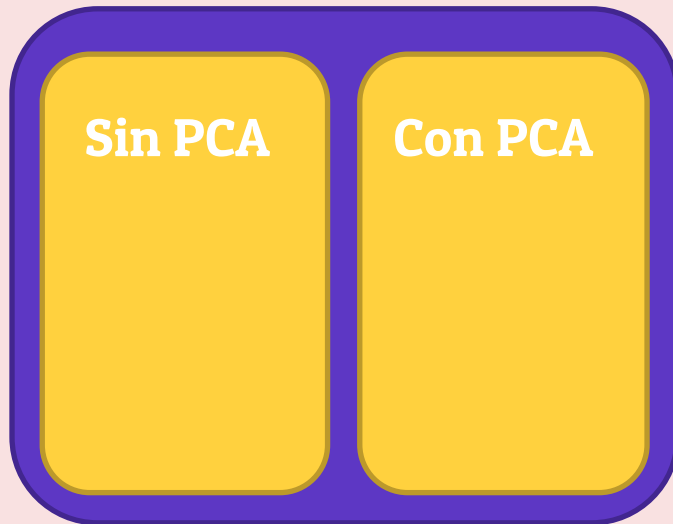
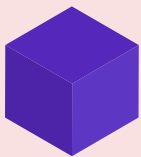


Figura 3. Metodología con Encoding y PCA .





03

Resultados



Resultados

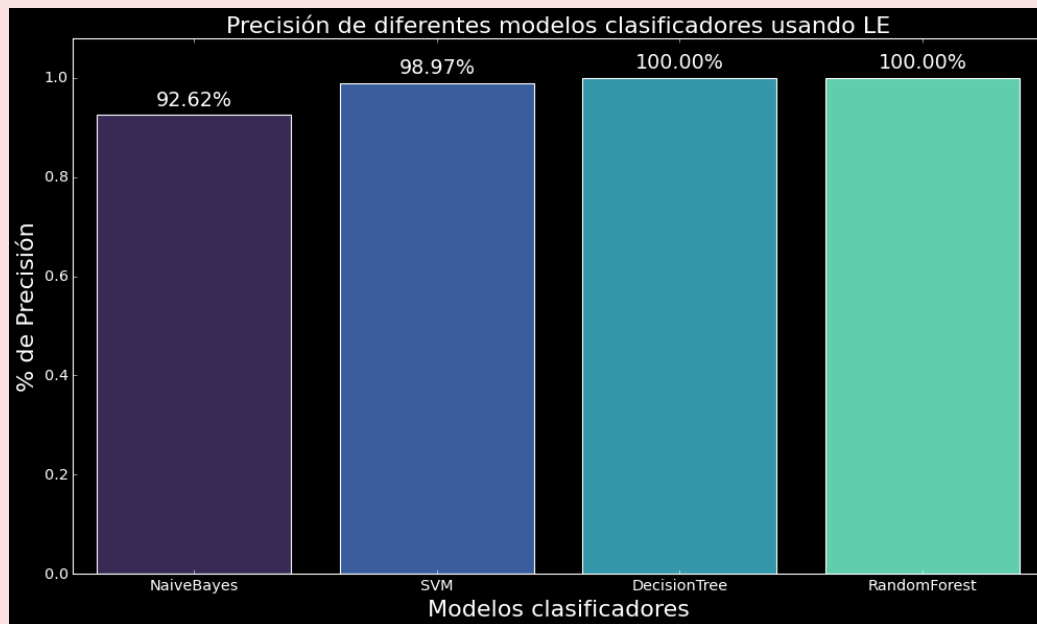


Gráfico 1. Clasificadores con Label Encoding

Resultados

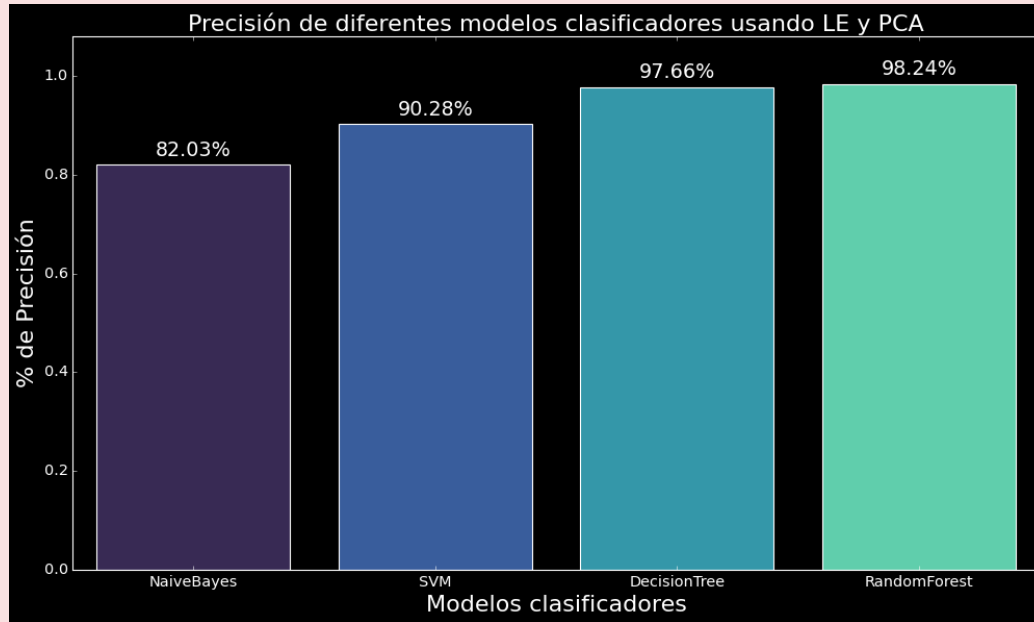


Gráfico 3. Clasificadores con Label Encoding y PCA.

Resultados

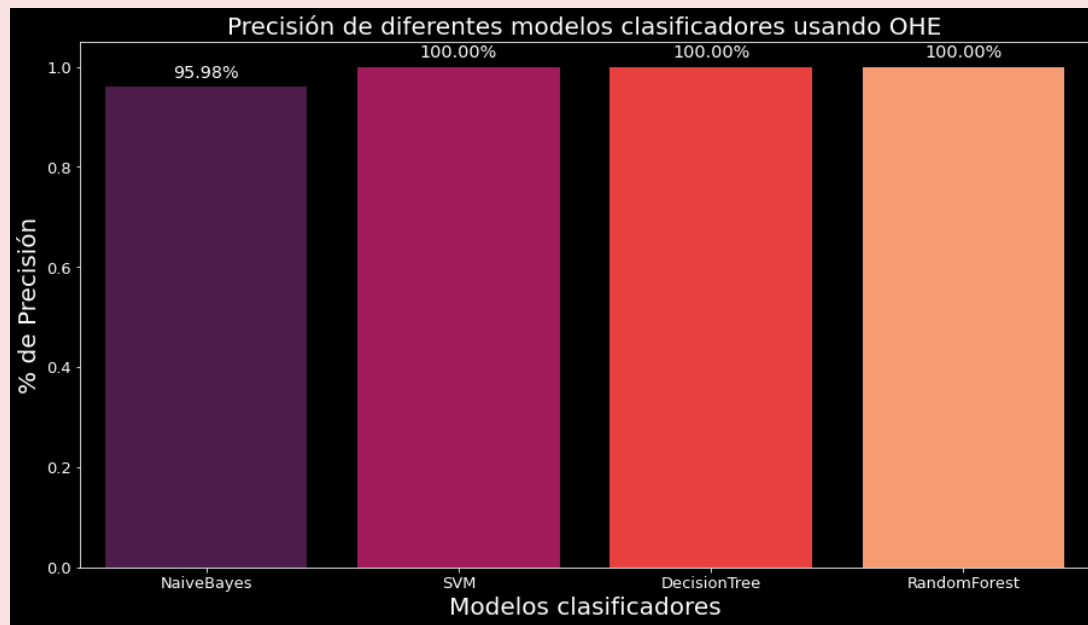


Gráfico 4. Clasificadores con One Hot Encoding

Resultados

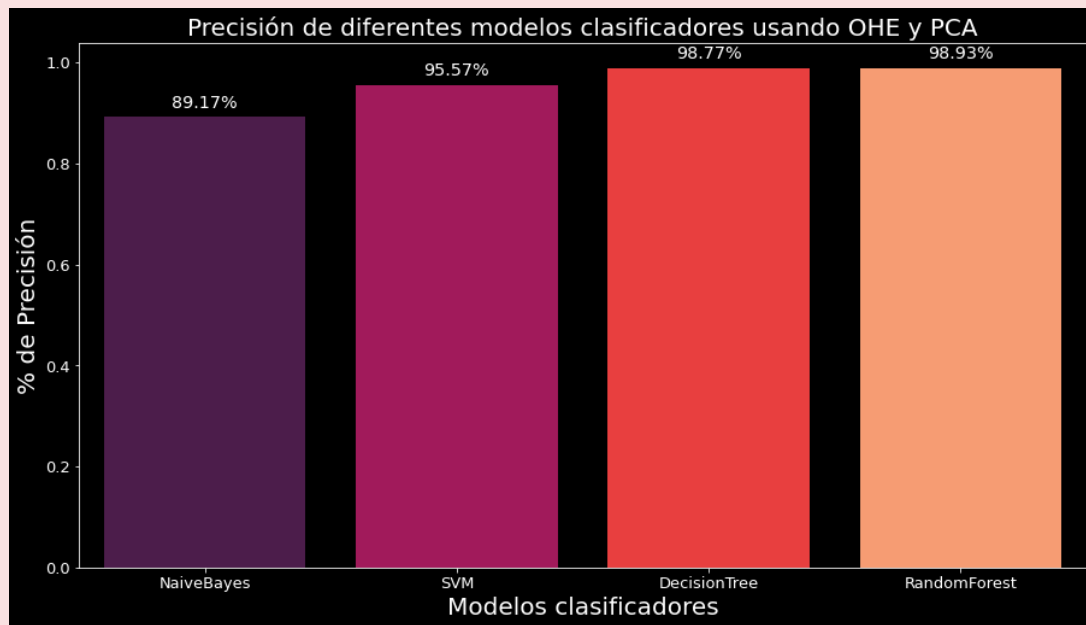


Gráfico 6. Clasificadores con One Hot Encoding y PCA

Resultados

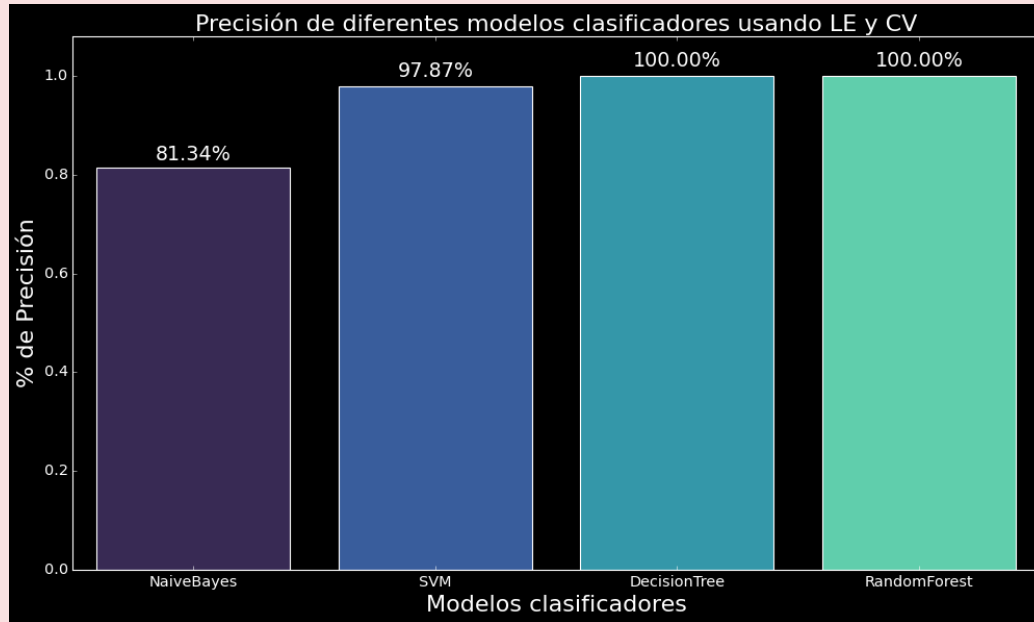


Gráfico 2. Clasificadores con Label Encoding y Cross Validation

Resultados

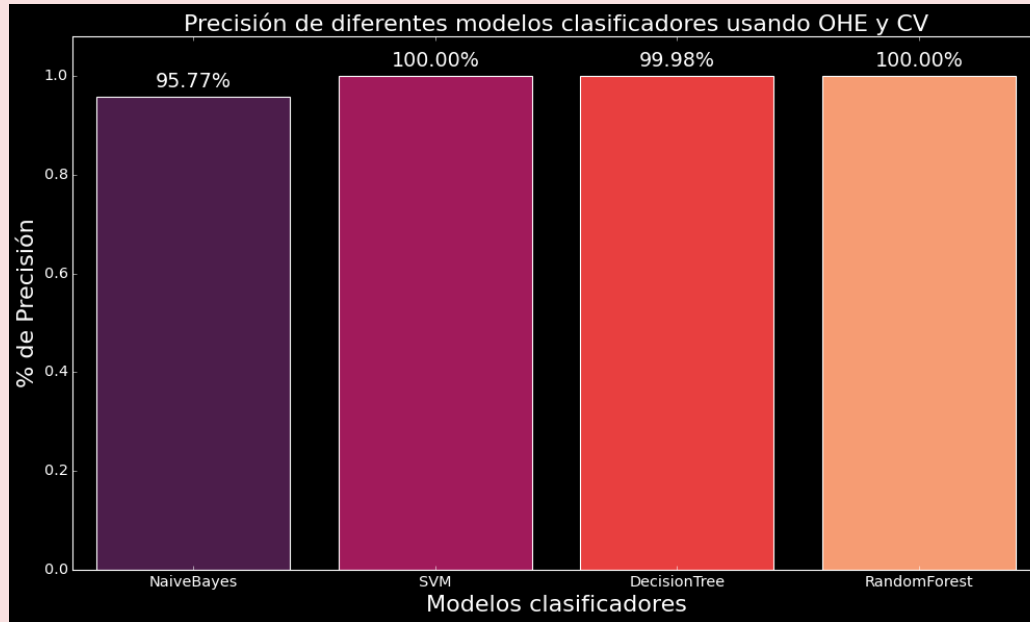
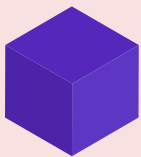


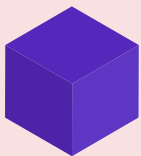
Gráfico 5. Clasificadores con One Hot Encoding y Cross Validation



04

Conclusiones





04

Conclusiones



- El Dataset utilizado para entrenar posee información de sobra para poder generar un modelo preciso.
- Es posible generar un modelo mas simple aplicando PCA sin perder mucha eficacia.
- Se generaron diversos modelos capaces de diferenciar correctamente entre una seta venenosa y una seta comestible.
- El clasificador Naive Bayes obtuvo los peores resultados para nuestro Dataset.



Gracias!



Systems
Fest
2021 • 1

Universidad
Industrial de
Santander

