



**UANL**

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

**FCFM**

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



# PRÁCTICA 3 CLASIFICACIÓN

## APRENDIZAJE AUTOMÁTICO



Profesor

JOSÉ ANASTACIO HERNÁNDEZ  
SALDAÑA

Alumna

JUDITH CAROLINA LUCIO  
GARZA

Matrícula: 1877415

NUESTRA BASE DE DATOS...  
PENGUINS

La base de datos de uso libre consta de registros de 3 distintas especies de pingüinos en lo que respecta al largo y profundidad de su pico, la longitud de sus aletas, así como su masa corporal. Los datos fueron recabados y publicados por el Dr. Kirsten Gorman y la Estación Palmer de la Antártica.

Durante esta práctica se tomarán en cuenta las variables referentes al pico de los pingüinos dependiendo de la especie a la que pertenezcan.  
A continuación se presentan las estadísticas descriptivas para cada una de ellas:

ADELIE			
Se tienen 146 ejemplares			
Longitud del Pico			
Promedio	38.82 mm	Minimo	32.1 mm
Máximo	46 mm	Desviación	2.66 mm
Profundidad del Pico			
Promedio	18.34 mm	Minimo	15.5 mm
Máximo	21.5 mm	Desviación	1.22 mm

# CHINSTRAP

Se tienen 68 ejemplares

Longitud del Pico			
Promedio	<b>48.83 mm</b>	Minimo	<b>40.9</b>
Máximo	<b>58 mm</b>	Desviación	<b>3.34 mm</b>

Profundidad del Pico			
Promedio	<b>18.42 mm</b>	Minimo	<b>16.4 mm</b>
Máximo	<b>20.8 mm</b>	Desviación	<b>1.14 mm</b>

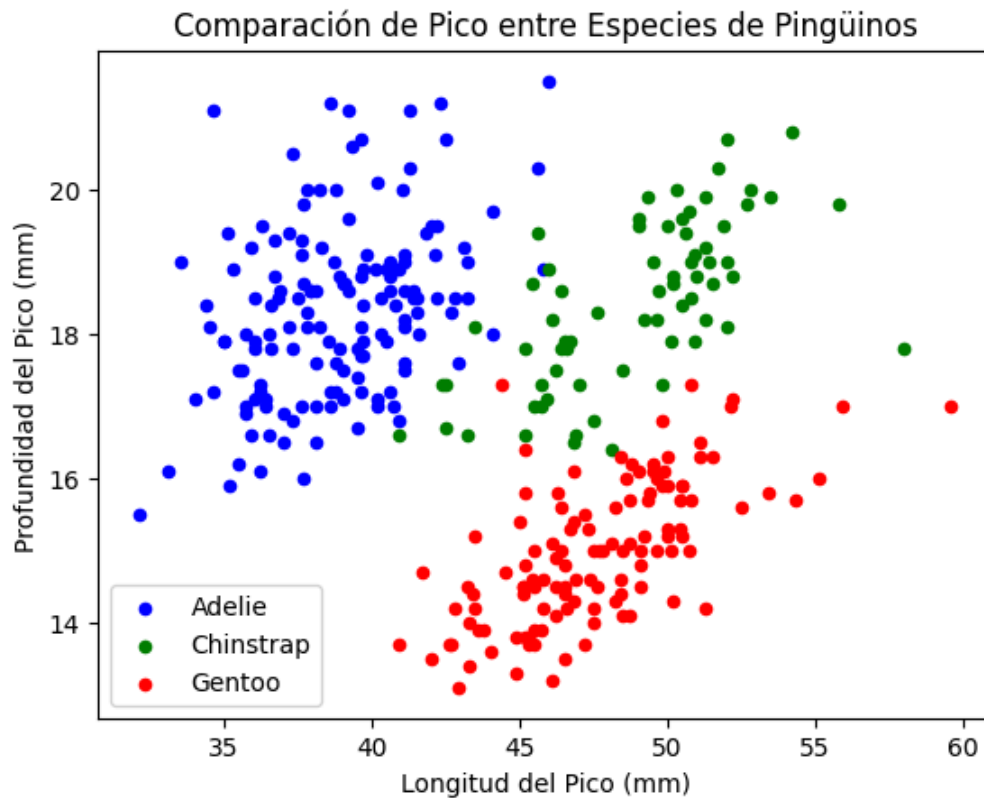
# GENTOO

Se tienen 119 ejemplares

Longitud del Pico			
Promedio	<b>47.57 mm</b>	Minimo	<b>40.9</b>
Máximo	<b>59.6 mm</b>	Desviación	<b>3.11 mm</b>

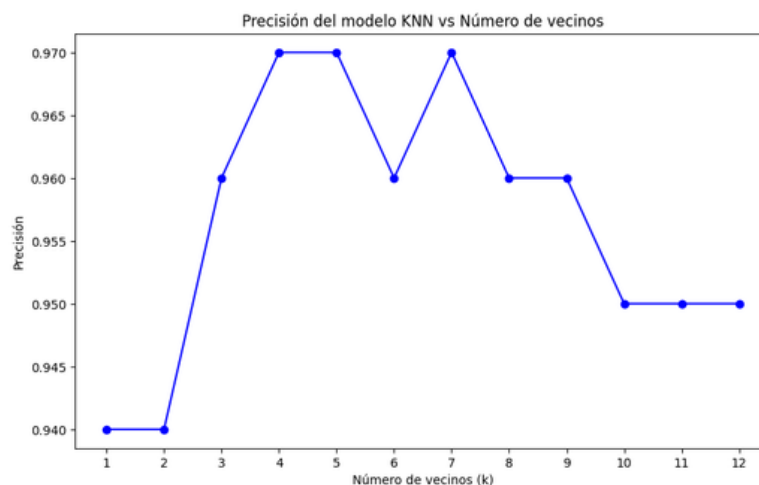
Profundidad del Pico			
Promedio	<b>14.99 mm</b>	Minimo	<b>13.1 mm</b>
Máximo	<b>17.3 mm</b>	Desviación	<b>0.99 mm</b>

Primeramente, se graficaron a manera de dispersión los registros correspondientes a cada especie, tomando en cuenta la longitud y profundidad de los picos. Se puede apreciar que se distribuyen de la siguiente manera:



Posteriormente, se tomaron 233 registros de pingüinos para conformar los datos de entrenamiento y 100 registros para usar como prueba con el método de K Nearest Neighbors.

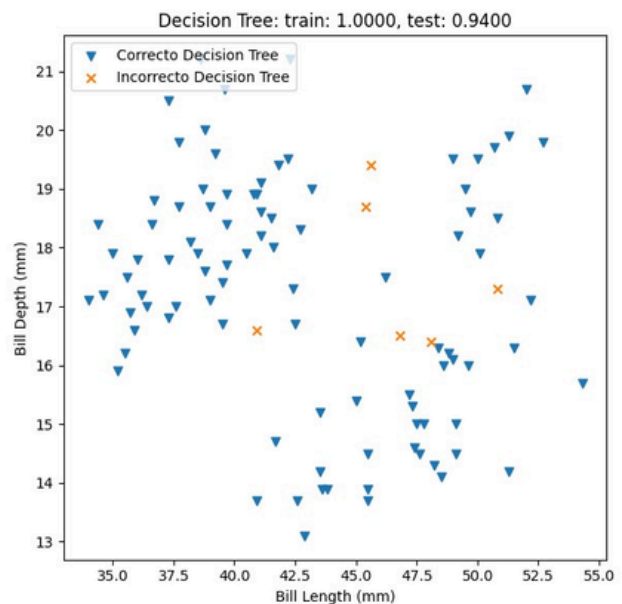
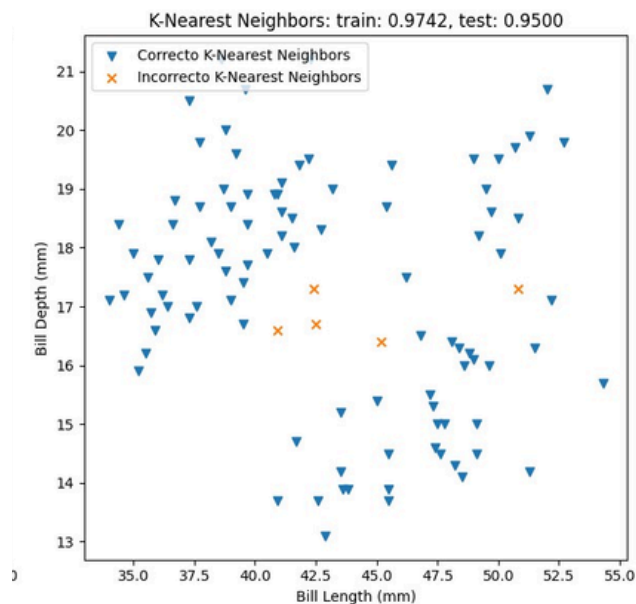
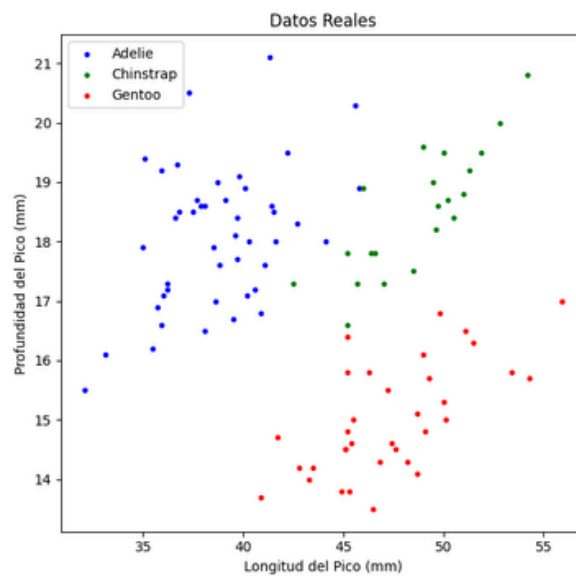
Se utilizó el parámetro  $k = 5$  dados los resultados de precisión observados en la gráfica siguiente.



La precisión del modelo KNN fue de 97% en entrenamiento y 95% en prueba.

Por otro lado, se utilizó también un Árbol de Decisión para clasificar a los pingüinos (diagrama anexo). Éste tuvo una precisión del 100% en entrenamiento y un 94% en prueba, lo que puede indicar overfitting.

El método elegido es el de **KNN**, debido a su ajuste en testing. A continuación se muestran gráficamente las distribuciones de los especímenes, así como los resultados de los dos métodos probados.



# ANEXO

Árbol de Decisión

