



<b>Actividad el algoritmo de bosque de decisión (Random Forest) (Tiempo estimado: 90 [minutos] ) Actividad individual.</b>	
<b>Objetivo de aprendizaje:</b>  Programar un algoritmo de predicción en múltiples variables en Python con el uso de las librerías de Scikit_Learn para crear un modelo de bosque de decisión.	<b>Habilidades:</b>  Razonamiento lógico y sistémico.
<b>Instrucciones:</b>  <ol style="list-style-type: none"><li>1. Revisa las <b>secciones 5 y 6</b> del siguiente link: <a href="https://www.kaggle.com/learn/intro-to-machine-learning">https://www.kaggle.com/learn/intro-to-machine-learning</a></li><li>2. Responde correctamente las siguientes preguntas guía: En la gráfica se muestran dos curvas de Sobreajuste/Sobreentrenamiento en aprendizaje supervisado. El error de entrenamiento se muestra en azul, mientras que el error de validación se muestra en rojo. Si el error de validación se incrementa mientras que el de entrenamiento decrece puede que se esté produciendo una situación de sobreajuste.</li></ol> <div data-bbox="568 934 1193 1375"></div> <ol style="list-style-type: none"><li>A) Define los términos Underfitting y overfitting, según se plantea en la sección 5.</li><li>B) En tus propias palabras, explica de manera sencilla, la lógica de la regresión Random Forest según se plantea en la sección 6.</li><li>C) ¿Qué ventaja tiene la regresión Random Forest en comparación con la regresión Decission Tree?</li></ol> <ol style="list-style-type: none"><li>3. En Google Colaboratory codifica el script que se adjunta a este documento.</li><li>4. ¿El error medio absoluto obtenido en la regresión Random Forest, es menor, mayor o igual que el error medio absoluto obtenido en la regresión DecissionTree? Justifica tu respuesta.</li><li>5. En un repositorio en GitHub agrega el script; y en los comentarios, las preguntas y respuestas.</li></ol>	
<b>Recomendaciones al facilitador:</b>  1. Descarga el archivo melb_data.csv	<b>Recursos y materiales necesarios:</b>  Cuenta de correo en Gmail Registro en Kaggle.com




# UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO

## Instalación Configuración y Comunicación de Sistemas Operativos

NOTA: el archivo original que puedes descargar en el minicurso, tiene más de 5000 registros; Google Colaboratory no permite cargar archivos con demasiados registros, es por ello que se redujo a los primeros 100.


2. El documento en el enlace <https://www.kaggle.com/learn/intro-to-machine-learning> es un minicurso, el cual pretende hacer una breve introducción al aprendizaje de máquinas. Es importante que leas con detenimiento las secciones 5 y 6.

 RandomForest.ipynb ☆  
Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda [Se guardaron todos los cambios](#)  
+ Código + Texto  
[17] 


```
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error
from sklearn.model_selection import train_test_split
import pandas as pd
```

  
[18] 


```
from google.colab import files
uploaded = files.upload()
```

  
 Seleccionar archivos melb\_data.csv  
• melb\_data.csv(text/csv) - 13839 bytes, last modified: n/a - 100% done  
Saving melb\_data.csv to melb\_data (1).csv  
  
[19] 

```
melbourne_data = pd.read_csv('melb_data.csv')
melbourne_data = melbourne_data.dropna(axis=0)
y=melbourne_data.Price
melbourne_features = ['Rooms', 'Bathroom', 'Landsize', 'Lattitude', 'Longtitude']
X = melbourne_data[melbourne_features]
```

```
train_X, val_X, train_y, val_y = train_test_split(X, y, random_state = 0)
melbourne_model = RandomForestRegressor(random_state=1)
melbourne_model.fit(train_X, train_y)
val_predictions = melbourne_model.predict(val_X)
print(mean_absolute_error(val_y, val_predictions))
```

  
 210016.0



UNIVERSIDAD AUTÓNOMA DEL ESTADO DE MÉXICO  
Instalación Configuración y Comunicación de Sistemas Operativos

RÚBRICA DE EVALUACIÓN DE LA INFOGRAFÍA				
PUNTO A EVALUAR	MUY BIEN 10	BIEN 8	REGULAR 5	PUNTAJE OBTENIDO
Información vertida en el documento	La información que se vierte en el documento es veraz y está en el contexto correcto.	La información que se vierte en el documento es veraz pero no está en el contexto correcto.	La información que se vierte en el documento no es veraz y no está en el contexto correcto.	
			TOTAL	