



**Tecnológico Nacional de México
Instituto Tecnológico de Culiacán**

Inteligencia Artificial

SCC1012_gA

Dr. José Mario Ríos Félix

Hora

06:00pm – 07:00pm

**Algoritmo SVM y KNN
Utilizando la DataSet “adult.cvs”**

Ingeniería en Sistemas Computacionales

Cejudo Rodríguez Julio Ángel

18170299

Culiacán, Sinaloa, México
Mayo del 2023

adult_csv.ipynb - Colaboratory

https://colab.research.google.com/drive/12c7PZQJySLgfM0eE264RI28Mu4DQFTN#scrollTo=rrW9_0P0QhHT

adult_csv.ipynb

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda

Comentario Compartir

Archivos

- drive
- sample_data
- adult_csv.csv

+ Código + Texto

```
[1] #Montar drive con google colab
from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

[2] #Importar la libreria pandas
import pandas as pd

[3] #Cargar el dataset
df = pd.read_csv('/content/adult_csv.csv')

#Mostrar los primeros 5 registros del dataset
df.head(5)
```

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capitalgain	capitalloss
0	2	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	1	
1	3	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	

Disco 84.43 GB de espacio disponible

completado a las 15:55

03:56 p. m. 01/06/2023

adult_csv.ipynb - Colaboratory

https://colab.research.google.com/drive/12c7PZQJySLgfM0eE264RI28Mu4DQFTN#scrollTo=rrW9_0P0QhHT

adult_csv.ipynb

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Se han guardado todos los cambios

Comentario Compartir

Archivos

- drive
- sample_data
- adult_csv.csv

+ Código + Texto

```
#Mostrar los primeros 5 registros del dataset
df.head(5)
```

	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capitalgain	capitalloss
0	2	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	1	
1	3	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	
2	2	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	
3	3	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	
4	1	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	

Disco 84.43 GB de espacio disponible

completado a las 15:55

03:56 p. m. 01/06/2023

adult_csv.ipynb - Colaboratory

https://colab.research.google.com/drive/12c7PZQJySLgfM0eE264RI28Mu4DQFTN#scrollTo=rrW9_0P0QhhT

adult_csv.ipynb

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Se han guardado todos los cambios

Archivos

- drive
- sample_data
- adult_csv.csv

+ Código + Texto

```
[5] #Modificar la tabla
X = df.iloc[:,1:-1]
X.head(5)
```

	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capitalgain	capitalloss	hours-per-week
0	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	1	0	40
1	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	45
2	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40
3	Private	234721	11th	7	Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40
4	Private	338409	Bachelors	13	Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40

0 s completado a las 15:55

03:57 p. m. 01/06/2023

adult_csv.ipynb - Colaboratory

https://colab.research.google.com/drive/12c7PZQJySLgfM0eE264RI28Mu4DQFTN#scrollTo=sPNUt07gRZHs

adult_csv.ipynb

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda

Archivos

- drive
- sample_data
- adult_csv.csv

+ Código + Texto

```
[6] #Revisar la info
X.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48842 entries, 0 to 48841
Data columns (total 13 columns):
#   Column             Non-Null Count  Dtype  
---  --
0   workclass          46043 non-null  object  
1   fnlwgt             48842 non-null  int64   
2   education          48842 non-null  object  
3   education-num      48842 non-null  int64   
4   marital-status     48842 non-null  object  
5   occupation         46033 non-null  object  
6   relationship       48842 non-null  object  
7   race               48842 non-null  object  
8   sex                48842 non-null  object  
9   capitalgain        48842 non-null  int64   
10  capitalloss        48842 non-null  int64   
11  hoursperweek       48842 non-null  int64   
12  native-country     47985 non-null  object  
dtypes: int64(5), object(8)
memory usage: 4.8+ MB
```

```
#Mostrar los primeros 5 registros de la columna "sex"
y = df['sex']
y.head(5)
```

0 s completado a las 15:55

03:58 p. m. 01/06/2023

adult_csv.ipynb - Colaboratory

https://colab.research.google.com/drive/12c7PZQCljySLgfM0eE264RI28Mu4DQFTN#scrollTo=sPNUt07gRZHs

adult_csv.ipynb

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Se han guardado todos los cambios

Archivos

- drive
- sample_data
- adult_csv.csv

+ Código + Texto

```
#Muestran los primeros 5 registros de la columna 'sex'
y = df['sex']
y.head(5)

0    Male
1    Male
2    Male
3    Male
4   Female
Name: sex, dtype: object

[8] #Importar librerías correspondientes para realizar el KNN y SVC (Utilizaremos la librería sklearn)
#Soporte de vectores "SVM"
from sklearn.svm import SVC
#Algoritmo de k vecinos mas cercanos "KNN"
from sklearn.neighbors import KNeighborsClassifier
#Hacer validaciones
from sklearn.model_selection import cross_val_score

[18] from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.svm import SVC
from sklearn.model_selection import cross_val_score

# Especifica las columnas categóricas que deseas codificar como una lista
categorical_cols = ['workclass', 'fnlwtg', 'education', 'education-num', 'marital-status', 'occupation', 'relationship', 'race', '

# Crea un objeto ColumnTransformer para aplicar la codificación one-hot
```

0 s completado a las 15:55

03:59 p. m. 01/06/2023

adult_csv.ipynb - Colaboratory

https://colab.research.google.com/drive/12c7PZQCljySLgfM0eE264RI28Mu4DQFTN#scrollTo=sPNUt07gRZHs

adult_csv.ipynb

Archivo Editar Ver Insertar Entorno de ejecución Herramientas Ayuda Se han guardado todos los cambios

Archivos

- drive
- sample_data
- adult_csv.csv

+ Código + Texto

```
[18] from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.svm import SVC
from sklearn.model_selection import cross_val_score

# Especifica las columnas categóricas que deseas codificar como una lista
categorical_cols = ['workclass', 'fnlwtg', 'education', 'education-num', 'marital-status', 'occupation', 'relationship', 'race',

# Crea un objeto ColumnTransformer para aplicar la codificación one-hot
preprocessor = ColumnTransformer(
    transformers=[('encoder', OneHotEncoder(), categorical_cols)],
    remainder='passthrough'
)

# Aplica la codificación one-hot a tus datos
X_encoded = preprocessor.fit_transform(X)

# Crea el modelo SVM y realiza la validación cruzada
svm = SVC()
svm_scores = cross_val_score(svm, X_encoded, y, cv=5, scoring='accuracy')

# Calcula el promedio de las corridas
svm_scores.mean()

1.0
```

0 s completado a las 15:55

03:59 p. m. 01/06/2023