



# ESCUELA POLITÉCNICA NACIONAL

## ESCUELA DE FORMACIÓN DE TECNÓLOGOS



### FUNDAMENTOS INTELIGENCIA ARTIFICIAL

ASIGNATURA:

Fundamentos de Inteligencia Artificial

PROFESOR:

Ing. Vanessa Guevara

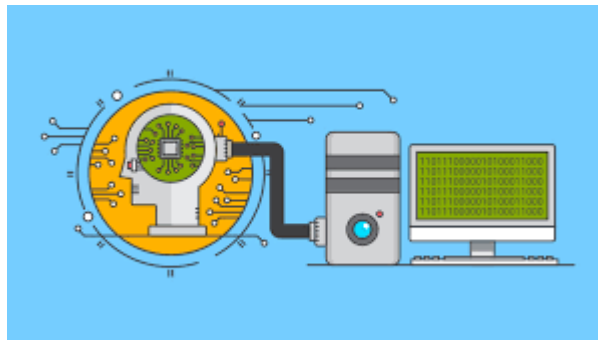
PERÍODO ACADÉMICO:

2024-A

## TAREA

TÍTULO:

**Aprendizaje Supervisado. Clasificación**



Vela David

## 1. PROPÓSITO DE LA PRÁCTICA

El propósito de esta tarea es construir un modelo de clasificación que pueda predecir el tipo de **medicamento a administrar a un paciente** en función de sus características personales.

Para ello, se proporciona un conjunto de datos que contiene información sobre diferentes individuos.

El conjunto de datos está en formato excel y contiene las siguientes columnas:

- age: edad del paciente
- sex: genero del paciente principal (0: masculino, 1: femenino)
- Blood Pressure Levels (BP): niveles de presión arterial (Bajo: 0, Normal: 1, Alto: 2)
- Cholesterol Levels: nivel de colesterol (Alto:1, Normal:0)
- Na to Potassium Ration: concentraciones de sodio y potasio en el cuerpo humano
- Drug: Tipos de medicamentos (0,1,2,3,4)

## 2. OBJETIVOS

- Familiarizarse con el proceso de resolución de un problema de clasificación utilizando aprendizaje supervisado.
- Entender el proceso para crear un modelo de clasificación y cómo evaluar su rendimiento utilizando métricas adecuadas.

## 3. INSTRUCCIONES

- Importar el conjunto de datos y realizar una exploración inicial para familiarizarse con la estructura y los tipos de datos.
- Dividir el conjunto de datos en un conjunto de entrenamiento y un conjunto de prueba. **(70% entrenamiento, 30% prueba)**
- Construir un modelo de clasificación utilizando el algoritmo de **árbol de decisión**.
- Entrenar el modelo utilizando el conjunto de entrenamiento y evaluar su rendimiento utilizando el conjunto de prueba.
- Evaluar las predicciones del modelo con las métricas de evaluación correspondiente
- Realizar predicciones del tipo de medicamento a administrar para un conjunto de datos de prueba adicional.

## 4. DESARROLLO DE LA PRÁCTICA

*Aquí detallar y explicar todas las etapas realizadas para construir el modelo. Incluir captura de pantallas.*

### 1. Importar datos

```
#Leer archivo - importar datos
import pandas as pd
datos= pd.read_excel("medicinas.xlsx")
print("Head with ID:\n")
print(datos.head())
print("Describe Data:\n")
print(datos.describe())
```

Con pandas leemos el archivo que es de tipo Excel e imprimimos los encabezados y su descripción

### 2. Dividir el conjunto de datos (70% entrenamiento, 30% prueba)

```
#Separar conjunto de características y etiquetas
x= datos.iloc[:,0:-1] #Características
y= datos.iloc[:,1] # Salidas todos excepto la última

#Dividir los datos en conjunto de entrenamiento y prueba
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x,y, test_size = 0.30, random_state = 70) #30% para prueba, semilla 70 (entrenamiento)
```

En este apartado lo que realizamos es separar el conjunto de datos en “x” y “y” y con la librería `train_test_split` dividimos el conjunto en un 30% para prueba y un 70 % para entrenamiento

### 3. Construir un modelo

```
#TRABAJAMOS CON ÁRBOL DE DECISIÓN
from sklearn.tree import DecisionTreeClassifier
modelo = DecisionTreeClassifier()
```

Trabajando con árbol de decisión creamos el modelo

### 4. Entrenar el modelo

```
#Entrenar modelo
modelo.fit(x_train,y_train)
```

Entrenamos el modelo

## 5. Métricas

```
# DESISIÓN - SÓLO PARA ARBOL DE DESISIÓN SE DETERMINA CUAL DE LAS CARACTERÍSTICAS ES MÁS IMPORTANTE SOLO ÁRBOL DE DESCIÓN
importancia = modelo.feature_importances_
print("Importancia de las características: \n")
print(importancia)

#Nombres de las características:
nombres= datos.columns.tolist()
for i,j in zip(importancia, nombres):
    print(j,i)

#Probar el modelo con los datos de prueba
y_obt= modelo.predict(x_test)

#Evaluar el modelo con las métricas
from sklearn.metrics import accuracy_score, confusion_matrix
exactitud = accuracy_score(y_test,y_obt)
print("\n")
print("La exactitud es: ", exactitud)
print("\n")
```

```
Importancia de las características:

[0.          0.14212828 0.          0.26583875 0.12159864 0.47043433]
Unnamed: 0 0.0
Age 0.14212827988338192
Sex 0.0
BP 0.265838753083651
Cholesterol 0.12159863945578232
Na_to_K 0.47043432757718473

La exactitud es: 1.0
```

Determinamos la importancia de cada una de las etiquetas para determinar cual es la más importante, además observamos la exactitud de 1, es decir, es un buen modelo ya que no se ha equivocado en ninguna predicción.

## 5. RESULTADOS

- Con el modelo desarrollado, predecir el tipo de medicamento a administrar para los siguientes casos:

```
#Realizar prediccion
# o Hombre, 45 años, con presión arterial alta, colesterol normal y concentración de sodio y potasio de 12.
tipo_medimento1 = modelo.predict([[45,0,2,0,12]])
#o Mujer, 30 años, con presión arterial normal, colesterol normal y concentración de sodio y potasio de 30.
tipo_medimento2 = modelo.predict([[30,1,1,1,30]])
#o Mujer, 65 años, con presión arterial bajo, colesterol alto y concentración de sodio y potasio de 7.
tipo_medimento3 = modelo.predict([[65,1,0,1,7]])
```

- Hombre, 45 años, con presión arterial alta, colesterol normal y concentración de sodio y potasio de 12.

```
El tipo_medimento es: [0]
*****
```

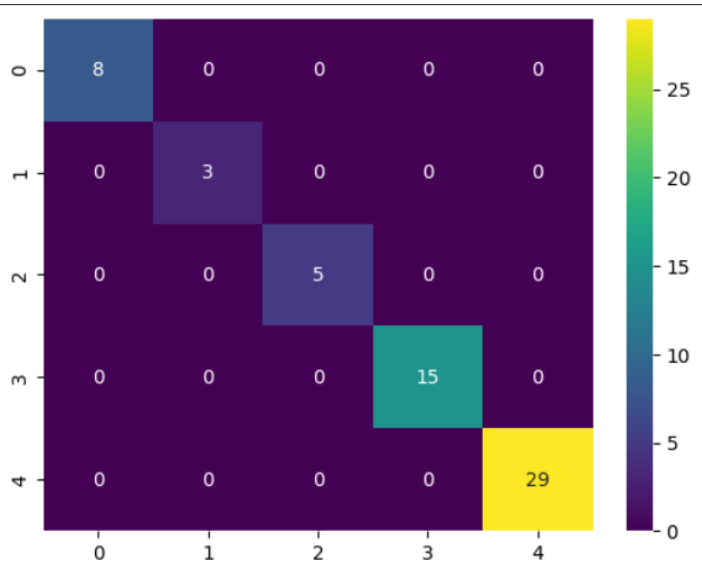
- Mujer, 30 años, con presión arterial normal, colesterol normal y concentración de sodio y potasio de 30.

```
El tipo_medimento es: [4]
*****
```

- Mujer, 65 años, con presión arterial bajo, colesterol alto y concentración de sodio y potasio de 7.

```
El tipo_medicamento es: [2]
*****
```

- Elaborar un mapa de calor para la matriz de confusión. Analizar e interpretar la matriz de confusión.



La diagonal nos indica los supuestos, falsos y verdaderos, en el grafico concluimos que su predicción ha sido excelente, no se ha equivocado.

- ¿Qué información le proporcionó la matriz de confusión sobre el rendimiento del modelo?

Que rinde muy bien, es decir, la matriz de confusión es un esquema que proporciona una visualización del rendimiento de un algoritmo de clasificación al mostrar las predicciones correctas e incorrectas de un modelo para cada etiqueta, y en nuestro caso predijo de manera correcta todos los supuestos, falsos y verdaderos.

- Explique si el modelo creado proporcionará una buena predicción sobre el tipo de medicamento a administrar.

Si, porque acertó todas las predicciones

- Indique las características del conjunto de datos que tienen mayor y menor importancia.

```
Importancia de las características:
[0.14212828 0.26583875 0.12159864 0.47043433]
Age 0.14212827988338192
Sex 0.0
BP 0.265838753083651
Cholesterol 0.12159863945578232
Na_to_K 0.47043432757718473
```

Las características más importantes son las de un número mayor, en este caso de más a menos importantes son:

Na\_to\_k > BP > Age > Cholesterol > Sex

## 6. CONCLUSIONES

- El modelo de árbol de decisión demostró alta precisión (100%) para predecir el tipo de medicamento.
- Las características más importantes fueron la relación de sodio/potasio, presión arterial y edad.
- La matriz de confusión y métricas como la exactitud permitieron evaluar y respaldar el buen rendimiento del modelo.