

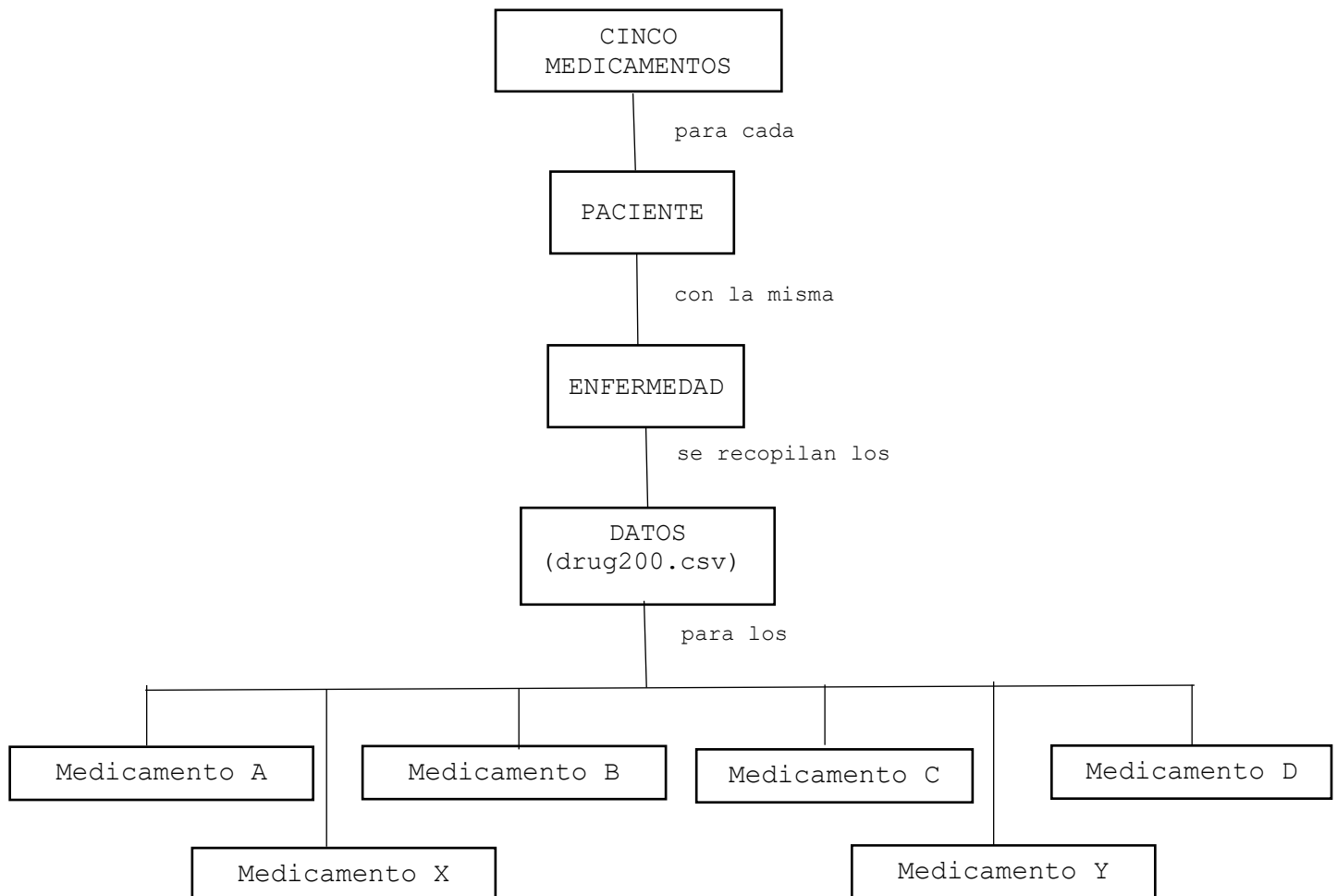
PREDECIR MEDICAMENTOS - PACIENTES

Imagina que eres un investigador médico que recopila datos para un estudio. Has recopilado datos sobre un conjunto de pacientes, todos los cuales padecían la misma enfermedad. Durante su curso de tratamiento, cada paciente respondió a uno de los 5 medicamentos: Medicamento A, Medicamento B, Medicamento C, Medicamento X e Y.

Parte de tu trabajo es construir un modelo para averiguar qué medicamento podría ser apropiado para un futuro paciente con la misma enfermedad. Las características de este conjunto de datos son la edad, el sexo, la presión arterial y el colesterol de los pacientes. El objetivo es identificar el fármaco al que respondió cada paciente.

Este es un ejemplo de un clasificador multiclase. Puedes utilizar la parte de entrenamiento del conjunto de datos para crear un árbol de decisión y luego utilizarlo para predecir la clase de un paciente desconocido o para recetar un medicamento a un nuevo paciente.

Fuente de DATOS: IBM



CIENCIA DE DATOS DE ENTRENAMIENTO DEL MODELO

Construir un modelo para averiguar qué medicamento podría ser apropiado para un futuro paciente con la misma enfermedad.

Para construir un modelo de ciencia de datos en Python que pueda recomendar un medicamento apropiado para un futuro paciente con una enfermedad específica, puedes seguir los siguientes pasos generales:

1. **Recopilación de datos:** Reúne un conjunto de datos que contenga información sobre diferentes medicamentos y los pacientes a los que se les han recetado. Los datos deben incluir características relevantes de los pacientes, como edad, género, síntomas, diagnóstico y medicamento recetado.
2. **Preprocesamiento de datos:** Realiza tareas de limpieza y preprocesamiento en tus datos. Esto puede incluir la eliminación de valores atípicos, manejo de valores faltantes, codificación de variables categóricas y normalización de características numéricas.
3. **Exploración y visualización de datos:** Realiza un análisis exploratorio de datos para obtener información sobre las características de los pacientes y los medicamentos recetados. Utiliza técnicas de visualización, como gráficos de barras, histogramas o diagramas de dispersión, para comprender mejor los patrones y las relaciones en los datos.
4. **División de datos:** Divide los datos en un conjunto de entrenamiento y un conjunto de prueba. El conjunto de entrenamiento se utilizará para construir el modelo, mientras que el conjunto de prueba se utilizará para evaluar su rendimiento.
5. **Selección y entrenamiento del modelo:** Selecciona un modelo de aprendizaje automático adecuado para tu problema. Puedes considerar el uso de algoritmos como árboles de decisión, bosques aleatorios, regresión logística o clasificadores de vectores de soporte. Entrena el modelo utilizando el conjunto de entrenamiento.
6. **Evaluación del modelo:** Evalúa el rendimiento del modelo utilizando el conjunto de prueba. Calcula métricas de evaluación, como precisión, recall, puntaje F1 o exactitud, para medir la calidad de las predicciones del modelo.
7. **Ajuste y optimización del modelo:** Si el rendimiento del modelo no es satisfactorio, puedes ajustar y optimizar los hiperparámetros del algoritmo de aprendizaje automático. Esto se puede hacer utilizando técnicas como la búsqueda en cuadrícula o la búsqueda aleatoria.
8. **Predicción de medicamentos para nuevos pacientes:** Una vez que estés satisfecho con el rendimiento del modelo, puedes utilizarlo para predecir qué medicamento podría ser apropiado para un nuevo paciente con una enfermedad específica. Asegúrate de preprocesar los datos del nuevo paciente de la misma manera que lo hiciste con los datos de entrenamiento.

Utilizando los datos que proporcionaste: Medicamento A, Medicamento B, Medicamento C, Medicamento X y Medicamento Y, con características como edad, sexo, presión arterial (BP), colesterol, relación de sodio a potasio (Na_to_K) y medicamento.

```

import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classification_report

# Paso 1: Recopilación de datos
# Supongamos que tienes un archivo CSV llamado 'drug200.csv' con las
# siguientes columnas: Age, Sex, BP, Cholesterol, Na_to_K, Drug.

# Paso 2: Carga y preprocesamiento de datos
data = pd.read_csv('drug200.csv')

# Paso 3: Exploración y visualización de datos
# Si es necesario, realiza un análisis exploratorio de datos y
# visualización para comprender los patrones y las relaciones en los
# datos.

# Paso 4: División de datos
X = data.drop('Drug', axis=1) # Características de los pacientes
y = data['Drug'] # Medicamentos recetados

# Codificar variables categóricas
le = LabelEncoder()
X['Sex'] = le.fit_transform(X['Sex'])
X['BP'] = le.fit_transform(X['BP'])
X['Cholesterol'] = le.fit_transform(X['Cholesterol'])

X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)

# Paso 5: Selección y entrenamiento del modelo
model = DecisionTreeClassifier()
model.fit(X_train, y_train)

# Paso 6: Evaluación del modelo
y_pred = model.predict(X_test)
print(classification_report(y_test, y_pred))

# Paso 7: Ajuste y optimización del modelo
# Si el rendimiento del modelo no es satisfactorio, puedes ajustar y
# optimizar los hiperparámetros del modelo DecisionTreeClassifier.

# Paso 8: Predicción de medicamentos para nuevos pacientes
# Supongamos que tienes un nuevo paciente con características dadas por
# una nueva fila de datos llamada 'nuevo_paciente'.
# Asegúrate de preprocesar los datos del nuevo paciente de la misma
# manera que los datos de entrenamiento.

nuevo_paciente = pd.DataFrame({'Age': [45], 'Sex': ['M'], 'BP':
['HIGH'], 'Cholesterol': ['HIGH'], 'Na_to_K': [0.9]})
nuevo_paciente['Sex'] = le.transform(nuevo_paciente['Sex'])
nuevo_paciente['BP'] = le.transform(nuevo_paciente['BP'])

```

```
nuevo_paciente['Cholesterol']  
le.transform(nuevo_paciente['Cholesterol'])  
  
medicamento_apropiado = model.predict(nuevo_paciente)  
print("Medicamento apropiado para el nuevo paciente:",  
medicamento_apropiado)
```

Se utiliza un modelo de árbol de decisión (DecisionTreeClassifier) para predecir el medicamento apropiado para un nuevo paciente en función de sus características. Las variables categóricas se codifican utilizando LabelEncoder y luego se divide el conjunto de datos en conjuntos de entrenamiento y prueba. Se entrena el modelo y se evalúa su rendimiento utilizando las métricas de clasificación. Finalmente, se realiza una predicción para un nuevo paciente y se muestra el medicamento apropiado.