

## **Pruebas realizadas al sistema supervisado de transporte masivo**

### **Capítulo 1**

#### **Introducción**

En este documento se describen las pruebas que se realizaron al sistema que se desarrolló para predecir rutas óptimas en un sistema de transporte masivo, utilizando métodos de aprendizaje supervisado. El modelo supervisado implementado es un árbol de decisión, el cual se entrenó con datos generados por un sistema basado en reglas.

El objetivo principal es demostrar la eficacia del aprendizaje supervisado en la clasificación de rutas como óptimas o no óptimas en función de características como las estaciones de origen y las de destino, el tiempo de viaje y las líneas utilizadas.

#### **Pruebas realizadas**

#### **Generación de datos**



Los datos que son necesarios para entrenar y evaluar el modelo supervisado fueron generados mediante un sistema basado en reglas. Este sistema considera conexiones directas e indirectas entre las estaciones y calcula el tiempo total de viaje para cada ruta.

#### **Resultado esperado**

Los archivos rutas\_generadas.csv y rutas\_generadas.xlsx contienen todas las rutas generadas, incluyendo las estaciones de origen, las de destino, el tiempo de viaje y si la ruta es óptima.

### **Método utilizado**

Los datos generados incluyen las etiquetas supervisadas como Ruta Óptima, que se clasifica como "Si" si el tiempo total es menor o igual a 60 minutos, y "No" en caso contrario. Estas etiquetas sirven como la variable objetivo (y) para el modelo supervisado.

### **Resultado esperado**

Los archivos rutas\_generadas.csv y rutas\_generadas.xlsx deben contener todas las rutas generadas, clasificadas como óptimas o no óptimas.

## Resultado

El archivo generado contiene las siguientes rutas:

Estación Origen	Estación Destino	Línea	Tiempo (min)	Ruta Óptima
Tatooine	Alderaan	Ruta 1	10	Si
Alderaan	Yavin IV	Ruta 2	20	Si
Yavin IV	Hoth	Ruta 3	30	Si
Hoth	Dagobah	Ruta 4	40	Si
Dagobah	Bespin	Ruta 5	50	Si
Bespin	Endor	Ruta 6	60	Si
Endor	Naboo	Ruta 7	70	No
Naboo	Coruscant	Ruta 8	80	No
Coruscant	Kamino	Ruta 9	90	No
Kamino	Mandalore	Ruta 10	100	No
Tatooine	Yavin IV	Ruta 1 + Ruta 2	30	Si
Alderaan	Hoth	Ruta 2 + Ruta 3	50	Si
Yavin IV	Dagobah	Ruta 3 + Ruta 4	70	No
Hoth	Bespin	Ruta 4 + Ruta 5	90	No
Dagobah	Endor	Ruta 5 + Ruta 6	110	No
Tatooine	Hoth	Ruta 1 + Ruta 2 + Ruta 3	60	Si
Alderaan	Dagobah	Ruta 2 + Ruta 3 + Ruta 4	90	No
Yavin IV	Bespin	Ruta 3 + Ruta 4 + Ruta 5	120	No
Tatooine	Dagobah	Ruta 1 + Ruta 2 + Ruta 3 + Ruta 4	100	No

## **Preparación de los datos**

Los datos generados fueron preparados para ser utilizados en el modelo supervisado. Esto incluyó:

La transformación de las columnas categóricas (Estación Origen, Estación Destino, Línea) en valores numéricos mediante la biblioteca LabelEncoder, la división de los datos en conjuntos de entrenamiento (80%) y prueba (20%) fueron generados utilizando la biblioteca train\_test\_split, estas dos bibliotecas pertenecen a la librería Scikit-learn.

## **Propósito**

La preparación de los datos es un paso fundamental en los métodos supervisados, ya que asegura que el modelo reciba las entradas numéricas que son compatibles y que su rendimiento pueda ser evaluado en datos no vistos.

## **Resultado esperado**

Los datos deben estar listos para entrenar y evaluar el modelo supervisado.

## **Resultado obtenido**

La transformación y división de los datos se realizó exitosamente.

Ejemplo de datos preparados

```
X (Características):  
[[0, 1, 0, 10],  
 [1, 2, 1, 20],  
 [0, 2, 0, 30]]  
  
y (Etiqueta objetivo):  
["Si", "Si", "Si"]
```

### Entrenamiento del modelo supervisado

Se evaluó el rendimiento del modelo utilizando el conjunto de datos de prueba., la métrica utilizada fue la precisión, que mide el porcentaje de predicciones correctas.

### Método utilizado

Se compararon las predicciones del modelo con las etiquetas reales del conjunto de prueba.

### Resultado esperado

Se espera obtener una alta precisión que indique que el modelo ha aprendido correctamente las reglas del sistema de transporte.

### Evaluación del modelo supervisado

Se evaluó el modelo utilizando los datos de prueba y se calculó la precisión.

### Resultado esperado

Una precisión alta indica que el modelo tiene un buen rendimiento.

### Resultado obtenido

Precisión del modelo: 1.0 (100%).

El modelo clasificó correctamente todas las rutas del conjunto de prueba.

### Ejemplo de predicción

Se probó el modelo con un ejemplo que fue tomado directamente del **Dataset** generado para predecir si una ruta es óptima.

### Entrada de prueba

#### Ejemplo:



```
X.iloc[0].values.reshape(1, -1)
prediccion = modelo.predecir(ejemplo)
print(f"Predicción para el ejemplo: {prediccion}")
```

### Resultado esperado

La predicción debe coincidir con la ruta óptima de la primera fila.

### Resultado obtenido

Etiqueta real: "Si".

Predicción del modelo: "Si".

El modelo predijo correctamente que la ruta es óptima.

## Capítulo 2

### Introducción

En este documento se describen las pruebas realizadas al sistema desarrollado para agrupar rutas en un sistema de transporte masivo utilizando métodos de **aprendizaje no supervisado**. El modelo no supervisado implementado es el algoritmo **K-Means**, el cual se aplicó a un dataset generado dinámicamente para identificar patrones y similitudes entre rutas.

El objetivo principal es demostrar la eficacia del aprendizaje no supervisado en la agrupación de rutas según sus características, como estaciones de origen y destino, tiempo de viaje y línea utilizada, con el fin de optimizar la planificación y el análisis de transporte.

### Generación de datos

Los datos fueron generados automáticamente a partir de un sistema basado en estaciones y sus conexiones. Estas conexiones incluyen las rutas directas entre estaciones, las líneas utilizadas y los tiempos de viaje.

### Método utilizado

Las conexiones entre las estaciones se definieron manualmente en el código, y los datos se estructuraron en un DataFrame que incluye las columnas: Estación Origen, Estación Destino, Línea y Tiempo (min).

## Resultado esperado

Los datos generados deben incluir todas las rutas entre estaciones definidas, con sus respectivas características.

Resultado obtenido:

El archivo generado contiene las siguientes rutas:

Estación Origen	Estación Destino	Línea	Tiempo (min)
Tatooine	Alderaan	Ruta 1	10
Alderaan	Yavin IV	Ruta 2	20
Yavin IV	Hoth	Ruta 3	30
Hoth	Dagobah	Ruta 4	40
Dagobah	Bespin	Ruta 5	50
Bespin	Endor	Ruta 6	60
Endor	Naboo	Ruta 7	70
Naboo	Coruscant	Ruta 8	80
Coruscant	Kamino	Ruta 9	90
Kamino	Mandalore	Ruta 10	100
Tatooine	Yavin IV	Ruta 1 + Ruta 2	30
Alderaan	Hoth	Ruta 2 + Ruta 3	50
Yavin IV	Dagobah	Ruta 3 + Ruta 4	70
Hoth	Bespin	Ruta 4 + Ruta 5	90
Dagobah	Endor	Ruta 5 + Ruta 6	110
Tatooine	Hoth	Ruta 1 + Ruta 2 + Ruta 3	60
Alderaan	Dagobah	Ruta 2 + Ruta 3 + Ruta 4	90
Yavin IV	Bespin	Ruta 3 + Ruta 4 + Ruta 5	120
Tatooine	Dagobah	Ruta 1 + Ruta 2 + Ruta 3 + Ruta 4	100

## Preparación de los datos

Los datos generados fueron preparados para su uso en el modelo de agrupamiento K-

Means. Esto incluyó:



### Transformación de datos categóricos:

Las columnas Estación Origen, Estación Destino y Línea (que contienen datos categóricos) fueron transformadas en valores numéricos utilizando la biblioteca LabelEncoder.

### Estandarización de los datos

Las características numéricas fueron escaladas utilizando StandardScaler para normalizar los datos. Esto asegura que todas las columnas tengan una media de 0 y una desviación estándar de 1, mejorando el rendimiento del algoritmo.

### Propósito

La preparación de los datos es un paso importante en el aprendizaje no supervisado, ya que garantiza que todas las características tengan una escala uniforme, evitando que una característica domine a las demás.

### Resultado esperado

Los datos deben estar listos para ser procesados por el modelo K-Means.

### Resultado obtenido

Los datos preparados presentan las siguientes características escaladas:

Estacion Origen	Estacion Destino	Linea	Tiempo (min)
-1.566.699	-1.566.699	-1.566.699	-1.566.699
-1.218.544	-1.218.544	-1.218.544	-1.218.544
-0.870388	-0.870388	-0.870388	-0.870388
-0.522233	-0.522233	-0.522233	-0.522233
-0.174078	-0.174078	-0.174078	-0.174078

## Aplicación del modelo K-Means

### Descripción

Se aplicó el algoritmo K-Means al dataset preparado para agrupar las rutas en clústeres.

Cada clúster representa un grupo de rutas con características similares.

### Método utilizado

Se utilizó el algoritmo K-Means de la biblioteca Scikit-learn con `n_clusters=3`.

El modelo asignó cada ruta a uno de los tres clústeres definidos.

Los clústeres se añadieron como una nueva columna al dataset (Cluster).

### Resultado esperado

Las rutas deben agruparse en tres clústeres según sus características (estaciones, tiempo de viaje y línea utilizada).

### Resultado obtenido

Las rutas fueron agrupadas de la siguiente manera:

Estacion Origen	Estacion Destino	Linea	Tiempo (min)	Cluster
Tatooine	Alderaan	Ruta 1	10	0
Alderaan	Yavin IV	Ruta 2	20	0
Yavin IV	Hoth	Ruta 3	30	1
Hoth	Dagobah	Ruta 4	40	1
Dagobah	Bespin	Ruta 5	50	2
Bespin	Endor	Ruta 6	60	2

## **Visualización de los resultados**

Se generó un gráfico de dispersión para visualizar cómo las rutas se agrupan en los clústeres identificados.

## **Método utilizado**

La visualización se realizó utilizando seaborn y matplotlib.

## **Ejes del gráfico**

X: Tiempo de viaje (Tiempo (min)).

Y: Estación de origen (Estacion Origen) codificada.

## **Resultado esperado**

Las rutas dentro de cada clúster deben agruparse visualmente en el gráfico.

## **Resultado obtenido**

El gráfico muestra cómo las rutas se distribuyen en tres clústeres distintos según sus características. Cada clúster está representado por un color diferente.

## **Conclusión**

El uso de los métodos supervisados y no supervisados en el análisis del transporte masivo demostró ser una solución robusta y efectiva. Mientras que el aprendizaje supervisado se utiliza para clasificar rutas específicas como óptimas o no, el aprendizaje no supervisado permite descubrir patrones y agrupaciones que pueden ser clave para mejorar la eficiencia del sistema.

Ambos enfoques proporcionaron información valiosa:

- El aprendizaje supervisado destacó por su precisión predictiva del 100%.
- El aprendizaje no supervisado permitió descubrir patrones y clústeres de rutas similares que no se habrían identificado fácilmente con análisis tradicionales.

Por tanto, es evidente que estos métodos de inteligencia artificial son herramientas poderosas para analizar y optimizar sistemas de transporte masivo, permitiendo una gestión más eficiente y una mejor experiencia para los usuarios.

