

Presentación del Análisis de Propiedades Inmobiliarias

1. Descripción del dataset y análisis exploratorio de datos.

El dataset utilizado contiene información sobre propiedades inmobiliarias que incluye las siguientes características:

- Fechas: fecha de inicio, fecha de fin y fecha de creación del registro de la propiedad.
- Ubicación: latitud, longitud y tres niveles de descripciones de la ubicación (como ciudad, barrio, etc.).
- Características: número total de habitaciones, número de dormitorios, número de baños, superficie total y superficie construida.
- Precio: precio de la propiedad y moneda utilizada.
- Descripción adicional: tipo de propiedad (casa, departamento, etc.) y el tipo de operación (venta, alquiler, etc.).

El análisis exploratorio de datos reveló que algunas propiedades tenían datos faltantes o valores no válidos (como precios o habitaciones en cero). Por esta razón, se aplicó un filtro para eliminar aquellos registros con precios o habitaciones igual a cero, lo que dejó un subconjunto de propiedades válidas para el análisis posterior.

2. Proceso de implementación del modelo de regresión lineal

El objetivo del análisis fue implementar un modelo de regresión lineal para estudiar la relación entre el número de habitaciones de una propiedad y su precio.

Pasos del proceso:

- Se dividió el dataset en dos conjuntos: un 75% para el entrenamiento del modelo y un 25% para pruebas.
- Se utilizó la técnica de regresión lineal simple, que ajusta una línea a los datos minimizando el error entre los valores reales (precios) y los valores estimados por la línea.

El modelo genera dos parámetros clave:

- Pendiente: mide cuánto cambia el precio en función del número de habitaciones.
- Intercepto: es el valor inicial del precio cuando el número de habitaciones es cero.

3. Resultados del modelo y análisis de su precisión.

Los resultados obtenidos en los conjuntos de entrenamiento y prueba fueron los siguientes:
Conjunto de Entrenamiento:

- Pendiente del modelo: Indica que, por cada habitación adicional, el precio de la propiedad aumenta proporcionalmente.
- Intercepto: Establece el precio de una propiedad sin habitaciones, lo cual es más un parámetro teórico.
- Error Cuadrático Medio (MSE): Mide el promedio de los errores al cuadrado entre los valores predichos y los valores reales. Un valor bajo indica que el modelo es

preciso.

- Coeficiente de Determinación (R^2): Este valor indica qué tan bien se ajustan los datos al modelo. Un valor de R^2 cercano a 1 significa que el modelo explica bien la variabilidad de los datos.

Conjunto de Prueba:

- MSE y R^2 también fueron calculados para el conjunto de prueba, que representa datos que no fueron usados para ajustar el modelo. Esto permite medir el rendimiento real del modelo en nuevos datos.

Los valores de R^2 fueron moderados, lo que indica que, si bien el modelo captura parte de la relación entre habitaciones y precio, hay otros factores que también influyen.

4. Conclusiones y posibles mejoras.

Conclusiones.

- El modelo de regresión lineal muestra que existe una correlación entre el número de habitaciones y el precio de las propiedades, pero no es el único factor determinante.
- El valor de R^2 en el conjunto de prueba sugiere que el modelo podría beneficiarse de la inclusión de otras variables, como la ubicación o la superficie construida, para mejorar su precisión.

Posibles mejoras:

- Añadir más características: Incluir más variables, como la ubicación o la superficie total, podría mejorar la capacidad del modelo para predecir el precio.
- Modelos más complejos: Se podrían explorar modelos más avanzados como la regresión polinomial o técnicas de machine learning para capturar relaciones no lineales.
- Regularización: Aplicar técnicas de regularización (como Lasso o Ridge) para evitar el sobreajuste y mejorar la generalización del modelo a nuevos datos.

Análisis del código

Este código realiza un análisis de propiedades inmobiliarias utilizando datos de un archivo CSV. Los principales pasos son cargar las propiedades, mostrar la información, y realizar un análisis estadístico con un modelo de regresión lineal. A continuación, se explica en detalle el código:

1. Inclusión de bibliotecas y estructuras

- `#include "AnalisisPropiedades.h"`: Se incluye un archivo de encabezado que define las estructuras `Propiedad`` y `RegresionLineal`` junto con las funciones

relacionadas.

- `vector<Propiedad>`: Utiliza un vector para almacenar objetos de tipo `Propiedad`, que es una estructura que contiene diversos atributos como la fecha, ubicación, número de habitaciones, precio, etc.

2. Función `cargarPropiedades`

Esta función carga los datos de un archivo CSV y los convierte en objetos `Propiedad`:

- Se abre el archivo con `ifstream`. Si no se puede abrir, se muestra un mensaje de error.
- Se leen los datos línea por línea utilizando un `stringstream` para descomponer cada línea según el formato del CSV.
- Los valores se asignan a un objeto de tipo `Propiedad`, y luego se agregan a un vector `propiedades`.

3. Función `mostrarPropiedades`

Esta función recorre el vector de propiedades y muestra los atributos principales (fechas, ubicación, habitaciones, precio) en la consola:

- Se utiliza un bucle `for` que recorre todas las propiedades y utiliza `cout` para imprimir los detalles.

4. Regresión lineal

- `generarModelo`: Esta función toma dos vectores de datos (`x` e `y`, en este caso el número de habitaciones y los precios) y calcula la pendiente y el intercepto de la regresión lineal utilizando las fórmulas de mínimos cuadrados.
- `estimarValor`: Calcula el valor estimado de `y` para un dado `x` usando el modelo generado por la regresión lineal.
- `calcularMSE`: Calcula el error cuadrático medio (MSE), que es una métrica para evaluar el rendimiento del modelo en cuanto a la diferencia entre los valores reales y los predichos.
- `calcularR2`: Calcula el coeficiente de determinación (R^2), que mide la proporción de la variabilidad en `y` explicada por el modelo.

5. `main`: Proceso principal

- Carga de propiedades: Se llama a la función `cargarPropiedades` para leer los datos del archivo CSV y almacenarlos en `listaPropiedades`.
- Filtrado de datos*: Se extraen las propiedades con un número válido de habitaciones y precios, y se almacenan en dos vectores: `listaHabitaciones` y `listaPrecios`.
- División de datos: Los datos se dividen en dos conjuntos: uno de entrenamiento (75%) y otro de prueba (25%) utilizando una distribución aleatoria.
- Ajuste del modelo: Se entrena un modelo de regresión lineal con los datos de

entrenamiento.

- Cálculo de métricas: Se calculan el MSE y el R^2 tanto para el conjunto de entrenamiento como para el de prueba.
- Guardado de resultados: Los resultados se escriben en un archivo de salida (` analisis_resultados.txt`).