

Laboratorio 8.

Puesta en producción de un modelo de ML

Contexto

El mercado inmobiliario en las grandes ciudades brasileñas ha experimentado cambios significativos en los últimos años debido a factores socioeconómicos y de desarrollo urbano. Para las agencias inmobiliarias e inversores, es crucial comprender los factores que influyen en el precio del alquiler de viviendas para tomar decisiones informadas y estratégicas.

Objetivo

Reforzar el aprendizaje en selección de modelos y despliegue de soluciones de Machine Learning utilizando un conjunto de datos real sobre alquiler de viviendas en Brasil. Este laboratorio permitirá a los estudiantes aplicar sus conocimientos teóricos en un escenario práctico y relevante para el mercado actual.

Parte 1: Selección de Modelos (40 puntos)

1.1. Conjunto de datos

Se proporcionará un conjunto de datos sobre alquileres de viviendas en Brasil en el 2020. Este dataset incluye las siguientes características:

- city: ciudad de ubicación
- area: área en metros cuadrados
- rooms: número de habitaciones
- bathroom: número de baños
- parking spaces: número de espacios para parqueo
- floor: nivel o piso en el que se encuentra
- animal: si se permiten animales o no
- furniture: si está amueblado o no
- hoa (R\$): impuesto de la Asociación de Dueños, en Reales Brasileños
- rent amount (R\$): mensualidad
- property tax (R\$): impuesto sobre bienes inmuebles
- fire insurance (R\$): seguro contra incendios
- total (R\$): total de la renta mensual

1.2. Proceso de selección de modelo

Realice los siguientes pasos para seleccionar un modelo final:

a) Exploración de datos (10 puntos):

- Realice un análisis exploratorio detallado del conjunto de datos.
- Identifique patrones, correlaciones y posibles outliers.
- Visualice la distribución de las variables clave.

b) Preprocesamiento y limpieza de datos (10 puntos):

- Maneje los valores faltantes y outliers de manera apropiada.
- Realice la codificación de variables categóricas.
- Normalice o estandarice las variables numéricas según sea necesario.

c) Entrenamiento y selección de modelos (15 puntos):

- Implemente al menos tres algoritmos diferentes (por ejemplo, Regresión Lineal, Random Forest, Gradient Boosting).
- Utilice técnicas de validación cruzada para evaluar el rendimiento de los modelos.
- Compare los modelos utilizando métricas como RMSE, MAE y R^2 .

d) Interpretación de resultados (5 puntos):

- Analice la importancia de las características en el modelo seleccionado.
- Explique las fortalezas y debilidades del modelo elegido.

Parte 2: Creación de un sitio web interactivo (55 puntos)

2.1. Diseño de la aplicación

Desarrolle una aplicación web interactiva que cumpla con los siguientes requisitos:

a) Funcionalidad para cargar datos (15 puntos):

- Implemente un formulario que permita a los usuarios ingresar los detalles de una propiedad.
- Valide los datos ingresados para asegurar su consistencia con el modelo.

b) Predicción y visualización (20 puntos):

- Utilice el modelo seleccionado para realizar predicciones del precio de alquiler.
- Muestre la predicción al usuario de manera clara y comprensible.
- Implemente visualizaciones interactivas que muestren:

- * La importancia de las características en el modelo.
- * Tendencias de alquiler en diferentes ciudades o barrios de Brasil.

c) Diseño y usabilidad de la interfaz (20 puntos):

- Cree una interfaz intuitiva y atractiva.
- Implemente mensajes de error claros y orientación al usuario.

Parte 3: Reflexión (5 puntos)

3.1. Informe de reflexión

Escriba un breve informe (máximo 2 páginas) donde reflexione sobre:

- Los desafíos encontrados durante el proceso de selección del modelo y su puesta en producción.
- Los aprendizajes más significativos obtenidos durante el desarrollo del laboratorio.
- Una evaluación crítica de las fortalezas y limitaciones del enfoque utilizado.
- Sugerencias para mejorar el proceso de desarrollo y despliegue de modelos de Machine Learning en el futuro.

Instrucciones adicionales

- El laboratorio se realizará en parejas para fomentar la colaboración y el intercambio de ideas.
- Se anima a los estudiantes a buscar y utilizar datos adicionales que puedan enriquecer el modelo, especialmente si tienen acceso a información específica sobre el mercado inmobiliario brasileño.

Rúbrica de Calificación (100 puntos)

1. Selección de Modelos (40 puntos)

- Exploración y visualización de datos: 10 puntos
- Preprocesamiento y limpieza de datos: 10 puntos
- Entrenamiento y selección de modelos: 15 puntos
- Interpretación de resultados y métricas: 5 puntos

2. Aplicación interactiva (55 puntos)

- Funcionalidad para cargar datos: 15 puntos
- Predicción y visualización de resultados: 20 puntos
- Diseño y usabilidad de la interfaz: 20 puntos

3. Reflexión (5 puntos)

- Profundidad y calidad de la reflexión: 5 puntos

Criterios de evaluación detallados

Criterio	Excelente	Satisfactorio	Necesita mejorar	Insuficiente
Exploración de datos	Análisis exhaustivo con visualizaciones claras y relevantes	Buen análisis con algunas visualizaciones útiles	Análisis básico con pocas visualizaciones	Análisis superficial sin visualizaciones
Preprocesamiento	Manejo óptimo de valores atípicos y codificación eficiente	Buen manejo de datos con algunas mejoras posibles	Preprocesamiento básico con errores menores	Preprocesamiento inadecuado o incompleto
Selección de modelos	Implementación y comparación detallada de múltiples modelos	Buena selección de modelos con comparación adecuada	Implementación básica de modelos con comparación limitada	Selección de modelos inadecuada o incompleta
Funcionalidad de la app	Carga de datos fluida y manejo de errores robusto	Buena funcionalidad con manejo de errores básico	Funcionalidad básica con algunos errores	Funcionalidad limitada o con errores graves
Visualizaciones	Visualizaciones interactivas y altamente informativas	Buenas visualizaciones con cierta interactividad	Visualizaciones básicas con poca interactividad	Visualizaciones pobres o ausentes
Diseño de interfaz	Interfaz intuitiva, atractiva	Buen diseño	Diseño básico con algunos problemas de usabilidad	Interfaz poco intuitiva
Reflexión	Análisis profundo y crítico con ideas originales	Buena reflexión con algunos puntos interesantes	Reflexión básica con poco análisis crítico	Reflexión superficial o irrelevante

1. Selección de Modelos (40 puntos en total)

- Exploración y preprocesamiento de datos (10 puntos):

* Excelente: 9-10 puntos

* Satisfactorio: 7.5-8.9 puntos

- * Necesita mejorar: 6-7.4 puntos
- * Insuficiente: 0-5.9 puntos
- **Entrenamiento y selección de modelos (20 puntos):**
 - * Excelente: 18-20 puntos
 - * Satisfactorio: 15-17.9 puntos
 - * Necesita mejorar: 12-14.9 puntos
 - * Insuficiente: 0-11.9 puntos
- **Interpretación de resultados y métricas (10 puntos):**
 - * Excelente: 9-10 puntos
 - * Satisfactorio: 7.5-8.9 puntos
 - * Necesita mejorar: 6-7.4 puntos
 - * Insuficiente: 0-5.9 puntos

2. Aplicación interactiva (55 puntos en total)

- **Funcionalidad para cargar datos y realizar predicciones (20 puntos):**
 - * Excelente: 18-20 puntos
 - * Satisfactorio: 15-17.9 puntos
 - * Necesita mejorar: 12-14.9 puntos
 - * Insuficiente: 0-11.9 puntos
- **Diseño y usabilidad de la interfaz (15 puntos):**
 - * Excelente: 13.5-15 puntos
 - * Satisfactorio: 11.25-13.4 puntos
 - * Necesita mejorar: 9-11.24 puntos
 - * Insuficiente: 0-8.9 puntos
- **Incorporación y claridad de visualizaciones (20 puntos):**
 - * Excelente: 18-20 puntos
 - * Satisfactorio: 15-17.9 puntos
 - * Necesita mejorar: 12-14.9 puntos
 - * Insuficiente: 0-11.9 puntos

3. Reflexión (5 puntos en total)

- **Profundidad y calidad de la reflexión (5 puntos):**
 - * Excelente: 4.5-5 puntos
 - * Satisfactorio: 3.75-4.49 puntos

* Necesita mejorar: 3-3.74 puntos

* Insuficiente: 0-2.99 puntos