

Proyecto para el módulo de Álgebra

Home Value Estimator

Instituto Superior Politécnico de Córdoba | Ciencia de datos e inteligencia artificial

Profesora: Sandra Aguirre

5 de Noviembre de 2024

Nombre del proyecto:

Home Value Estimator: predictor de precios de viviendas mediante machine learning, este proyecto propone el desarrollo de un modelo de machine learning específicamente de regresión lineal múltiple que implementa algoritmos para analizar datos y proporcionar un predictor de precios según características de las viviendas.

Tipo de proyecto

Tecnológico: El proyecto involucra la creación de una aplicación de análisis de datos, la implementación de modelos de machine learning y la investigación en la efectividad de estos modelos.

Espacios participantes en el módulo

Álgebra

Estos ejes temáticos se relacionan con los siguientes espacios curriculares del módulo:

1. Ciencia de datos:
 - Programación: Conocimientos de lenguajes de programación como Python.
 - Algoritmos y estructuras de datos para diseñar y optimizar algoritmos de IA.
 - Inteligencia artificial y aprendizaje automático para fundamentos y técnicas de IA Y ML.
2. Matemáticas y Estadística:
 - Cálculo: para entender los modelos matemáticos detrás de los algoritmos de IA.
 - Álgebra lineal: fundamental para el funcionamiento de algoritmos de ML.d
 - Estadística: Para el análisis de datos y la validación de modelos

Competencias y habilidades del perfil profesional

1. Pensamiento crítico y resolución de problemas:

- Descripción: Análisis de datos complejos para tomar decisiones informadas, identificación de patrones y tendencias de los datos.
- Competencias: La capacidad de evaluar y sintetizar información y aplicar el razonamiento lógico para resolver problemas.

2. Habilidades técnicas:

- Descripción: Conceptos y aplicaciones del álgebra computacional, dominio de Python, herramientas como Matplotlib, Seaborn, Pandas, Scikit-learn.
- Competencias: Desarrollo de habilidades en programación, análisis de datos y uso de herramientas avanzadas de visualización, machine learning y álgebra computacional.

3. Trabajo en equipo y comunicación:

- Descripción: Colaboración con el equipo de proyecto y comunicación clara de resultados.
- Competencias: Desarrollo de habilidades de colaboración y trabajo en equipo, esenciales para trabajar de manera multidisciplinaria.

Problemas o necesidades:

En el mercado inmobiliario actual, la evaluación precisa de los precios de las propiedades es crucial tanto para compradores como vendedores, sin embargo, determinar el valor de una casa puede ser un proceso complejo y subjetivo, influenciado por múltiples factores como el número de habitaciones, baños, tamaño de propiedad, entre otros.

La falta de una herramienta confiable y automatizada para predecir los precios puede llevar a estimaciones inexactas, afectando las decisiones de compra y venta, y generando insatisfacción entre las partes involucradas en el proceso.

Fundamentación

Este proyecto fue elegido debido a la facilidad de acceso a datos de precios de viviendas, lo que permite probar y validar el funcionamiento de la predicción de precios de casas usando regresión múltiple.

El potencial de este proyecto es significativo, ya que puede transformar la manera en que se realizan las evaluaciones inmobiliarias, con este proyecto se pueden considerar múltiples factores que afecten el precio de la vivienda, proporcionando estimaciones más precisas y confiables. Esto beneficiaría a agentes inmobiliarios, tasadores, compradores y vendedores.

Con respecto a los profesionales de ciencia de datos este proyecto es altamente relevante, ya que nos permite abordar técnicas avanzadas de modelado, análisis de datos y desempeño. Además que proporciona experiencia práctica en el uso de herramientas necesarias para la cursada.

Visión del proyecto:

El proyecto tiene como objetivo desarrollar un modelo de regresión lineal múltiple para predecir los precios de casas basándose en sus características. Se estima que el modelo proporcionará una herramienta precisa y eficiente para evaluar el valor de las propiedades inmobiliarias.

- **Objetivo general:** Desarrollar un modelo de regresión lineal múltiple que prediga con precisión los precios basándose en las características específicas relevantes según el análisis.
- **Objetivos específicos:**
 1. Recolectar y preparar datos: Obtener un conjunto de datos de precios de viviendas y sus características, y realizar la limpieza y normalización de los datos.

2. Desarrollar el modelo: Implementar el modelo de regresión lineal múltiple utilizando técnicas de aprendizaje automático.
 3. Evaluar el modelo: Validar el modelo utilizando métricas de rendimiento como el R^2 , el error cuadrático medio (MSE) y la precisión.
 4. Predicción: Aplicar el modelo para predecir precios de viviendas.
 5. Integración de Conceptos de Álgebra y Matemática: Utilizar herramientas matemáticas para el análisis de datos.
- Metas: Calidad: Alcanzar una precisión del modelo de al menos 80% en la predicción de precios de viviendas, Cantidad: Evaluar el modelo utilizando al menos 1000 registros de datos de viviendas.

Selección de acciones

Objetivo específico	Acciones	Habilidades a lograr
Recolectar datos y limpieza	Buscar fuentes de datos confiables, limpiar y normalizar los datos.	Análisis de datos, manipulación de datos, limpieza de datos.
Desarrollar el modelo	implementar el modelo de regresión lineal múltiple utilizando Python y bibliotecas como scikit-learn.	Programación en Python, modelado predictivo, uso de bibliotecas de aprendizaje automático.

Evaluar el modelo	Validar el modelo utilizando un conjunto de datos de prueba y calcular métricas de rendimiento.	Evaluación de modelos, interpretación de métricas de rendimiento.
Predicción	Utilizar el modelo seleccionado para predecir precios en nuevos datos, interpretar los resultados y verificar su efectividad.	Habilidad para aplicar e interpretar modelos predictivos a datos reales.
Integración de Conceptos de Álgebra	Utilizar conceptos estadísticos y matemáticos para la interpretación de resultados y validación del modelo	Dominio de álgebra lineal y estadística aplicada. habilidad para integrar estos conceptos matemáticos.

Producto final:

1. Análisis del Problema

El mercado inmobiliario actual requiere herramientas precisas y eficientes para evaluar el valor de las propiedades. La falta de métodos automatizados y confiables puede llevar a estimaciones inexactas, afectando las decisiones de compra y venta. Este proyecto aborda esta necesidad desarrollando un modelo de regresión lineal múltiple para predecir los precios de las casas basándose en sus características.

2. Diseño del Modelo

El modelo de regresión lineal múltiple se diseñó para predecir los precios de las casas utilizando variables independientes como el número de habitaciones, baños, tamaño del sótano, espacio habitable y otras características relevantes obtenidas según los análisis.

3. Desarrollo del Prototipo y resultado

El prototipo del modelo se desarrolló en Python utilizando bibliotecas como scikit-learn para la implementación del modelo de regresión lineal múltiple. El proyecto ha demostrado ser una herramienta útil para evaluar el valor de las propiedades inmobiliarias, pero necesita más precisión.

Resultados del modelo desarrollado:

El modelo de regresión lineal múltiple desarrollado para predecir los precios de las casas utilizando las características de las propiedades ha mostrado resultados moderadamente satisfactorios. A continuación, se presentan las conclusiones clave basadas en las métricas de evaluación:

Coefficiente de Determinación R^2 , Valor: 0.51

Interpretación: Aproximadamente el 51% de la variabilidad en los precios de las casas puede ser explicada por las variables independientes seleccionadas (habitaciones, baños, tamaño del sótano, espacio interior sobre el nivel del suelo y espacio habitable). Esto indica que el modelo tiene una capacidad explicativa moderada, pero aún hay un 49% de la variabilidad que no está siendo capturada, lo que sugiere la presencia de otros factores influyentes no considerados en el modelo.

Error Cuadrático Medio (MSE), Valor: 74,198,772,469.47

Interpretación: El MSE indica que, en promedio, las predicciones del modelo se desvían de los valores reales en una cantidad significativa. Un MSE alto sugiere que las predicciones del modelo no son muy precisas y que hay margen para mejorar la precisión del modelo.

Observaciones Adicionales

Relación Positiva: Las variables seleccionadas (habitaciones, baños, tamaño del sótano, espacio interior sobre el nivel del suelo y espacio habitable) muestran una relación positiva con el precio de las casas. A medida que aumentan estas características, también lo hace el valor de la propiedad.

Variabilidad No Capturada: La recta de regresión no captura por completo la variabilidad en los datos, lo que indica que hay otros factores importantes que afectan el precio de las casas y que no están incluidos en el modelo actual.

Recomendaciones para Mejorar el Modelo

Incluir más variables: Considerar otras características que puedan influir en el precio de las casas, como la ubicación, el estado de la propiedad, las tendencias del mercado y las características del vecindario.

Utilizar modelos más complejos: Explorar el uso de modelos de aprendizaje automático más avanzados, como árboles de decisión, bosques aleatorios o redes neuronales, que puedan capturar relaciones más complejas entre las variables.

Validación y Optimización Continua: Realizar validaciones cruzadas y ajustes de hiper parámetros para optimizar el rendimiento del modelo y reducir el error de predicción.

En resumen, aunque el modelo actual proporciona una base sólida para la predicción de precios de viviendas, hay oportunidades significativas para mejorar su precisión y capacidad explicativa mediante la inclusión de más variables y el uso de técnicas de modelado más avanzadas.