

Análisis de datos y predicción mediante Machine Learning aplicado a Transacciones de propiedades

ACKERLEY FABRIZIO RUMICHE HIDALGO, CHRISTIAN HUMBERTO MONTENEGRO OLIVA, LEONEL EMILIANO HERNANDEZ LAVADO, JEREMY ALEXANDER AQUINO GIRON and Estrada-Rayme, Leighton

Leandro, M.Sc.⁶

^{1,2,3,4,5}Universidad Privada del Norte, Perú, n00197072@upn.pe, n00225371@upn.pe, n00271409@upn.pe, n00225564@upn.pe, leighton.estrada@upn.pe

SISTEMAS INTELIGENTES Y MACHINE LEARNING

Resumen—

El objetivo del proyecto actual es analizar y predecir el comportamiento de las transacciones inmobiliarias mediante el uso de técnicas de aprendizaje automático. Se implementarán modelos de regresión múltiple para identificar patrones y tendencias en el mercado de bienes raíces utilizando un conjunto de más de 10.000 registros y 12 características. Para garantizar la calidad del análisis, los datos se limpiarán. Después, los modelos predictivos se ajustarán y los datos se dividirán en conjuntos de entrenamiento y prueba. Además, se analizarán los coeficientes de regresión para comprender los elementos que tienen el mayor impacto en los precios de las propiedades. Este estudio mejorará la comprensión del mercado inmobiliario y ayudará a los inversores y agentes inmobiliarios a tomar decisiones. Este informe incluye el proceso de implementación de los modelos en un entorno de desarrollo Colaboratory y Kaggle para facilitar la reproducción y validación de los resultados. Por último, pero no menos importante, el uso de técnicas de aprendizaje automático en el análisis de transacciones inmobiliarias ofrece una herramienta poderosa para predecir precios y comprender las dinámicas del mercado como en nuestro proyecto.

El resumen debe contener entre 200 a 250 palabras [ES AL FINAL DEL TRABAJO]

Palabras clave—Lista de “palabras clave” separado por comas

Machine Learning, Predicción, Regresión Múltiple, Transacciones Inmobiliarias, Análisis de Datos

Ejemplo: Machine Learning, DataSet, Scikit-Learn

Abstract—

El resumen traducido a ingles

Keywords-- List at most 5 key index terms here.

I. INTRODUCCION

Considerar en este apartado mínimamente: Motivación del proyecto, estudios relacionados al problema que se esta trabajando, importancia del proyecto, importancia del ML actual en sus diversas aplicaciones.

Motivación del proyecto

La creciente demanda de análisis y predicción en el sector inmobiliario ha impulsado el uso de técnicas de Machine Learning para obtener resultados más precisos y eficientes. Este proyecto busca implementar un sistema de predicción de precios de propiedades utilizando diversas técnicas de Machine Learning para mejorar la toma de decisiones en la compra y venta de inmuebles.

Estudios Relacionados:

Importancia del proyecto

Este proyecto es importante porque proporciona una herramienta basada en IA que puede analizar grandes volúmenes de datos y generar predicciones precisas, lo cual es crucial en el mercado inmobiliario. El uso de Machine Learning en este contexto permite identificar patrones y tendencias que no son evidentes mediante métodos tradicionales.

Importancia del Machine Learning

El Machine Learning ha demostrado ser una herramienta poderosa en diversos campos, desde la medicina hasta el comercio, debido a su capacidad para procesar grandes volúmenes de datos y generar modelos predictivos precisos. Su aplicación en la predicción de transacciones de propiedades representa una innovadora forma de mejorar la eficiencia y efectividad en la toma de decisiones.

1.1. ANTECEDENTES

(Mínimo 9 antecedentes)

Ejemplo de cómo redactar un antecedente:

La investigación de Lopez (2019), tuvo por objetivo el desarrollo de una app móvil para mejorar el índice de adopción de perros en la ciudad de Trujillo, para lo cual utilizó el marco de trabajo ágil scrum, logrando como resultado aumentar en 10% el índice de adopciones en la ciudad de Trujillo.

1. MARTÍNEZ ET AL. (2020) PREDICIERON EL VALOR DE MERCADO DE LAS PROPIEDADES EN LIMA CON UNA PRECISIÓN DEL 85 % UTILIZANDO TÉCNICAS DE REGRESIÓN LINEAL.

2.EL ESTUDIO DE FLORES Y RAMÍREZ (2019) UTILIZÓ MÉTODOS DE AGRUPAMIENTO PARA DIVIDIR EL MERCADO INMOBILIARIO Y OFRECER RECOMENDACIONES PERSONALIZADAS A LOS COMPRADORES.

3. RODRÍGUEZ ET AL. (2020) EXPLORARON CÓMO LOS ELEMENTOS SOCIOECONÓMICOS PUEDEN AFECTAR LA DEMANDA DE PROPIEDADES EN ÁREAS URBANAS MEDIANTE EL USO DE MODELOS DE REGRESIÓN MÚLTIPLE.

4. LA INVESTIGACIÓN DE GUPTA Y MEHTA (2021) UTILIZÓ MODELOS DE REGRESIÓN LOGÍSTICA MÚLTIPLE PARA PREDECIR LOS PRECIOS DE ALQUILER EN GRANDES CIUDADES, LO QUE RESULTÓ EN UNA REDUCCIÓN DEL ERROR DE PREDICCIÓN DEL 12%.

5. EL ESTUDIO DE SMITH Y JOHNSON (2018) UTILIZÓ MODELOS DE ÁRBOLES DE DECISIÓN PARA CLASIFICAR PROPIEDADES SEGÚN SU PROBABILIDAD DE VENTA RÁPIDA. EL RESULTADO FUE UNA REDUCCIÓN DEL 15% EN EL TIEMPO PROMEDIO DE VENTA.

6. SANTOS Y FERNÁNDEZ (2022) UTILIZARON UN SISTEMA QUE UTILIZA ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA PREDECIR LOS CAMBIOS EN LOS PRECIOS DE LAS PROPIEDADES COMERCIALES EN LIMA, LO QUE LES PERMITIÓ OBTENER UNA PRECISIÓN DEL 10% EN COMPARACIÓN CON MODELOS CONVENCIONALES

7. ZHAO Y YANG (2021) UTILIZARON TÉCNICAS DE MACHINE LEARNING PARA PREDECIR LOS PRECIOS DE LAS VIVIENDAS EN EL MERCADO INMOBILIARIO DE BEIJING. EL ESTUDIO DEMOSTRÓ QUE LOS MODELOS DE MACHINE LEARNING PUEDEN PROPORCIONAR PREDICCIONES MÁS PRECISAS QUE LOS MÉTODOS TRADICIONALES.

8. LI Y WU (2020) APLICARON ALGORITMOS DE MACHINE LEARNING COMO EL BOSQUE ALEATORIO Y EL SOPORTE VECTORIAL PARA PREDECIR LOS PRECIOS DE LAS PROPIEDADES EN SHENZHEN. LOS RESULTADOS INDICARON QUE ESTOS MODELOS SON EFECTIVOS PARA CAPTAR LAS TENDENCIAS DEL MERCADO Y OFRECER PREDICCIONES PRECISAS.

9. AHMED Y MOUSTAFA (2022) INTEGRARON ANÁLISIS DE BIG DATA CON TÉCNICAS DE MACHINE LEARNING PARA ANALIZAR Y PREDECIR TRANSACCIONES DE PROPIEDADES EN NUEVA YORK. LOS MODELOS DESARROLLADOS AYUDARON A IDENTIFICAR PATRONES Y TENDENCIAS CLAVE EN EL MERCADO INMOBILIARIO, UTILIZANDO GRANDES VOLÚMENES DE DATOS SOCIOECONÓMICOS Y DEMOGRÁFICOS.

1.2. PLANTEAMIENTO DEL PROBLEMA

Detallar la situación actual de la problemática encontrada, en específico del campo de acción donde se enfocará el proyecto.

El mercado inmobiliario enfrenta desafíos significativos en la valoración precisa de propiedades debido a la dependencia de métodos tradicionales subjetivos, la diversidad y calidad variable de los datos provenientes de múltiples fuentes, y la influencia de factores económicos y sociales. Los métodos actuales son ineficientes y propensos a errores, complicando la toma de decisiones en la compra-venta, alquiler e inversión de inmuebles. La integración de técnicas de Machine Learning podría revolucionar este sector, proporcionando modelos predictivos más precisos y objetivos basados en datos históricos y actuales. Sin embargo, la implementación de ML presenta desafíos técnicos, como la selección de algoritmos adecuados, la interpretación de modelos complejos, y la escalabilidad, así como consideraciones éticas y legales relacionadas con la privacidad y equidad. Este proyecto se enfocará en desarrollar e implementar modelos de ML que mejoren la precisión en la valoración de propiedades y predigan tendencias del mercado, asegurando el cumplimiento de normativas de protección de datos y evitando sesgos discriminatorios, para crear un mercado inmobiliario más eficiente y transparente.

1.3. OBJETIVOS

1.3.1 Objetivo general

Desarrollar un sistema de predicción de precios de propiedades utilizando técnicas de Machine Learning para mejorar la toma de decisiones en el mercado inmobiliario.

1.3.2. Objetivos específicos

Recolectar y preprocesar datos relevantes del mercado inmobiliario:

- Identificar y obtener datasets de propiedades incluyendo características como ubicación, tamaño, tipo de propiedad, y precios históricos.
- Limpiar y normalizar los datos para asegurar su calidad y consistencia.

Seleccionar y entrenar modelos de Machine Learning adecuados:

- Implementar algoritmos de regresión lineal, árboles de decisión y redes neuronales.
- Entrenar los modelos con los datos preprocesados y ajustar hiperparámetros para optimizar el rendimiento.

Evaluar el desempeño de los modelos:

- Realizar pruebas de validación cruzada para medir la precisión y robustez de los modelos.
- Comparar los resultados obtenidos con los precios reales para evaluar la exactitud de las predicciones.

Desarrollar una interfaz de usuario para visualizar las predicciones:

- Crear una aplicación web que permita a los usuarios ingresar datos de propiedades y obtener predicciones de precios.
- Implementar gráficos y visualizaciones para facilitar la interpretación de los resultados.

Analizar los resultados y validar las predicciones:

- Comparar las predicciones generadas por el sistema con los datos del mercado real para identificar patrones y tendencias.
- Realizar ajustes en los modelos según sea necesario para mejorar la precisión y utilidad de las predicciones.

II. MARCO TEÓRICO

1.4. HERRAMIENTAS DE INGENIERÍA

Python: Python es un lenguaje de programación de alto nivel y propósito general que se ha convertido en una herramienta fundamental en el ámbito del Machine Learning y la ciencia de datos. Sus características clave incluyen una sintaxis clara y legible, y una extensa colección de bibliotecas y frameworks especializados. Según van Rossum y Drake (2009), Python es altamente valorado por su versatilidad y capacidad para facilitar el desarrollo rápido de aplicaciones complejas .

Scikit-Learn: Es una biblioteca de Python que proporciona herramientas simples y eficientes para el análisis de datos y la construcción de modelos predictivos. Está diseñada para interoperar con otras bibliotecas de Python como NumPy y SciPy, lo que la hace muy flexible y potente para tareas de Machine Learning. Pedregosa et al. (2011) destacan que Scikit-Learn es conocida por su facilidad de uso, robustez y documentación exhaustiva, lo que la convierte en una opción preferida tanto para principiantes como para expertos en el campo .

TensorFlow: Es una plataforma de código abierto desarrollada por Google Brain para la investigación y la producción en Machine Learning y redes neuronales profundas. TensorFlow permite a los desarrolladores construir y desplegar modelos de aprendizaje automático en una variedad de dispositivos y entornos. Abadi et al. (2016) mencionan que su diseño modular y escalable lo hace adecuado para una amplia gama de aplicaciones, desde investigación académica hasta sistemas de producción a gran escala .

Robustez y Capacidad para Manejar Grandes Conjuntos de Datos: La robustez se refiere a la capacidad de una herramienta o sistema para manejar errores, condiciones inesperadas y perturbaciones sin fallar o degradar significativamente su rendimiento. En el contexto del Machine Learning, esto implica que las herramientas puedan gestionar datos ruidosos, incompletos y de gran volumen de manera eficiente (Zhang et al., 2017).

Seleccione herramientas y/o técnicas de IA apropiadas para el desarrollo de la solución a un problema complejo de ingeniería, reconociendo las limitaciones y ventajas y justifique su selección.

1.5. ALGORITMOS

Se utilizarán algoritmos de regresión lineal, árboles de decisión y redes neuronales para la predicción de precios de propiedades. Estos algoritmos se seleccionaron por su capacidad para manejar diferentes tipos de datos y su eficacia probada en tareas de predicción.

Descripción de los algoritmos de IA a utilizar en el proyecto, incluir los modelos matemáticos usados en estos algoritmos como regresión lineal, árboles de búsqueda, etc. Incluir el pseudocódigo.

III. METODOLOGÍA

2.1. MATERIALES Y MÉTODOS UTILIZADOS

Hardware: Computadora con capacidad para ejecutar google Colaboratory y bibliotecas de Python.

Software: Entorno de desarrollo en Kaggle, Colaboratory, bibliotecas de Python:

*Selección de las herramientas hardware y software, **fundamentado***

2.2. PROCESO DE DESARROLLO DEL PROYECTO

Describe las fases de acuerdo con la metodología seleccionada.

Se debe adjuntar las evidencias (algoritmos, formular, fotografías de prototipos, etc)

IV. RESULTADOS

4.1. JUSTIFICACIÓN DEL PROYECTO

Justificar la aplicación de LOS ALGORITMOS en base a las pruebas de rendimiento, justificación teórica, experimentación y validación.

4.2. RESULTADOS OBTENIDOS DEL PROYECTO EN BASE A LOS OBJETIVOS

RESULTADO OE1: Definir la situación actual del problema

RESULTADO OE2: Diseño e implementación de la solución

Incluir el bloque del código del algoritmo más importante que evidencie el algoritmo de la IA declarado en el marco teórico.

RESULTADO OE3: Analizar los resultados obtenidos de la implementación

RESULTADO OG: Utilizar algoritmos de IA para identificar patrones, tendencias y áreas de mejora

V. DISCUSIÓN Y CONCLUSIONES

4.1. DISCUSIÓN DE RESULTADOS CON LOS ANTECEDENTES

4.2. CONCLUSIONES POR OBJETIVO

VI. REFERENCIAS BIBLIOGRÁFICAS

Mínimo 9 citas y referencias de artículos científicos o trabajos de investigación según el formato APA 7ma edición.

Flores, D., & Ramírez, H. (2019). Segmentación del mercado inmobiliario mediante técnicas de clustering. *Data Science Journal*, 200-215.

Martínez, A., et al. (2020). Aplicación de técnicas de regresión para la predicción de valores inmobiliarios. *Journal of Real Estate Research*, 45-60.

Smith, R., & Johnson, P. (2018). Clasificación de propiedades para venta rápida mediante árboles de decisión. *Real Estate Analytics*, 35(4), 112-125.

Rodríguez, F., et al. (2020). Análisis de factores socioeconómicos en la demanda de propiedades urbanas. *Urban Studies*, 50(2), 78-89.

Zhao, L., & Yang, Y. (2021). Machine learning techniques for real estate price prediction: A case study of Beijing housing market. *Expert Systems with Applications*, 183, 115452.

Li, X., & Wu, Q. (2020). Predicting real estate prices using machine learning algorithms: A study of Shenzhen housing market. *Journal of Property Research*, 37(3), 267-290.

Ahmed, S., & Moustafa, M. (2022). Integrating big data analytics and machine learning for real estate market analysis: Evidence from New York City. *International Journal of Information Management*, 62, 102437.

VII. ANEXOS

ANEXO 01: Fotografías de prototipos, algoritmos (código), producto mínimo viable (MVP), etc (resultado final)