

Proyecto de Integración y Automatización de Datos para IA para el proyecto XXX

Carolina Tobaría¹, Andrés Bravo², Juan Morales³

¹⁻²Dpto. de Ingeniería, ³Dpto. de Comunicación,

Universidad Central

Maestría Analítica de Datos

Curso de Integración y Automatización de Datos para IA

Bogotá, Colombia

¹ctobariac@ucentral.edu.co, ²abravop@ucentral.edu.co, ³jmoralesr10@ucentral.edu.co

April 6, 2024

Contents

1	Introducción (Max 250 Palabras) - (<i>Primera entrega</i>)	3
2	Características del proyecto de investigación que hace uso de Integración y Automatización de Datos para IA (Max 500 Palabras) - (<i>Primera entrega</i>)	3
2.1	Titulo del proyecto de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	3
2.2	Objetivo general (Max 100 Palabras) - (<i>Primera entrega</i>)	3
2.2.1	Objetivos específicos (Max 100 Palabras) - (<i>Primera entrega</i>)	4
2.3	Alcance (Max 200 Palabras) - (<i>Primera entrega</i>)	4
2.4	Pregunta de investigación (Max 100 Palabras) - (<i>Primera entrega</i>)	5
2.5	Hipotesis (Max 100 Palabras) - (<i>Primera entrega</i>)	5
3	Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (<i>Primera entrega</i>)	6
3.1	¿Cuál es el origen de los datos e información ? (Max 100 Palabras) - (<i>Primera entrega</i>)	6
3.2	¿Cuáles son las consideraciones legales o éticas del uso de la información? (Max 100 Palabras) - (<i>Primera entrega</i>)	6
3.3	¿Cuáles son los retos de la información y los datos que utilizara en Integración y Automatización de Datos para IA? (Max 100 Palabras) - (<i>Primera entrega</i>)	6

3.4	¿Que espera de la utilización de Integración y Automatización de Datos para IA para su proyecto? (Max 100 Palabras) - (<i>Primera entrega</i>)	7
4	Diseño de integración y Automatización de Datos para IA (Diagrama) (<i>Primera entrega</i>)	8
5	integración de Datos (<i>Segunda entrega</i>)	9
6	Automatización de Datos (<i>Segunda entrega</i>)	9
7	IA (<i>Segunda entrega</i>)	9
8	Proximos pasos (<i>Tercera entrega</i>)	10
9	Lecciones aprendidas (<i>Tercera entrega</i>)	11
10	Bibliografía	12

1 Introducción (Max 250 Palabras) - (*Primera entrega*)

En las grandes ciudades, la búsqueda de una vivienda adecuada a un precio justo se ha convertido en un verdadero reto. Bogotá, como capital y una de las ciudades más pobladas de Colombia, no es la excepción. El mercado inmobiliario de arrendamientos en la ciudad es altamente dinámico y complejo, donde arrendatarios y arrendadores enfrentan incertidumbres respecto al valor justo. Esta variabilidad en los precios obedece a múltiples factores como ubicación, características de la propiedad, oferta, demanda y cambios demográficos, dificultando establecer precios justos y competitivos.

En este contexto surge la necesidad de herramientas analíticas que permitan comprender los patrones y tendencias que rigen los precios de arriendo en las diferentes localidades de Bogotá. El análisis de datos y la inteligencia artificial emergen como soluciones prometedoras, al procesar grandes volúmenes de información y extraer tendencias valiosas para tal fin.

Una fuente clave de datos es *Finca Raíz*, principal plataforma online para el mercado inmobiliario colombiano [?]. Su API abierta brinda acceso a una extensa base de propiedades en arriendo, con detalles de ubicación, características y precios. Aprovechar esta valiosa fuente permite obtener una visión integral y actualizada del mercado de arrendamientos en Bogotá.

Este trabajo desarrolla un modelo predictivo basado en aprendizaje automático que utiliza los datos de Finca Raíz para estimar con precisión los precios de arriendo en las distintas localidades de Bogotá. Mediante el análisis de datos históricos de arrendamientos, características de propiedades y factores socioeconómicos, se busca brindar una herramienta útil para inquilinos, propietarios e inversionistas inmobiliarios.

Este enfoque analítico optimiza la toma de decisiones, proporcionando estimaciones confiables de precios justos y tendencias futuras, contribuyendo a mayor transparencia y eficiencia en el sector inmobiliario.

2 Características del proyecto de investigación que hace uso de Integración y Automatización de Datos para IA (Max 500 Palabras) - (*Primera entrega*)

2.1 Título del proyecto de investigación (Max 100 Palabras) - (*Primera entrega*)

Análisis y Predicción de Precios de Arriendo en las Localidades de Bogotá

2.2 Objetivo general (Max 100 Palabras) - (*Primera entrega*)

Desarrollar un sistema automatizado que, mediante técnicas de aprendizaje automático y análisis de datos obtenidos de la plataforma Finca Raíz, prediga los precios de arriendo en Bogotá y visualice los resultados a través de un tablero interactivo en Power BI y una base de datos SQL, facilitando así la toma de decisiones informadas en el mercado inmobiliario.

2.2.1 Objetivos específicos (Max 100 Palabras) - (*Primera entrega*)

- Implementar un proceso de automatización para la recolección y preprocesamiento semanal de datos de arriendo desde la plataforma Finca Raíz.
- Analizar los factores determinantes de los precios de arriendo en Bogotá.
- Establecer una base de datos Oracle.
- Desarrollar un modelo predictivo utilizando técnicas de aprendizaje automático, que sea capaz de estimar los precios de arriendo en las distintas localidades de Bogotá basándose en los factores identificados.
- Crear un tablero en Power BI para la visualización de los resultados del análisis y las predicciones de precios de arriendo en las localidades de Bogotá.

2.3 Alcance (Max 200 Palabras) - (*Primera entrega*)

Este proyecto abarca varias fases cruciales en el procesamiento de datos y la inteligencia de negocios, con un enfoque particular en el mercado inmobiliario de Bogotá, el proyecto se propone seguir la siguiente estructura de componentes principales:

- **Extracción de datos:** implementar un sistema automatizado que a través de consumo de API en la Web fincaraiz.com.co extraiga semanalmente la información de los arriendos en Bogotá.
- **Integración de datos:** Crear y mantener un proceso ETL que garantice que los datos extraídos se limpien, transformen y carguen adecuadamente en una base de datos Oracle para su almacenamiento y consulta eficiente.
- **Análisis de datos:** Aplicar métodos estadísticos y de machine learning para analizar los datos recopilados, lo que incluirá el desarrollo de un modelo de regresión lineal para predecir precios de arriendo en base a diferentes características de las propiedades en la ciudad de Bogotá.
- **Desarrollo de Modelos Predictivos:** Construcción de un modelo predictivo de precios de arriendo, utilizando técnicas de aprendizaje automático para identificar tendencias y patrones dentro de los datos históricos.
- **Visualización y Reporte:** Diseñar un dashboard interactivo usando Power BI que permita a los usuarios finales visualizar los datos y las predicciones de una manera comprensible y accionable.
- **Retroalimentación y Mejora Continua:** Establecer un mecanismo de feedback que permita la mejora continua del sistema basándose en el rendimiento del modelo predictivo y las necesidades cambiantes del negocio o del mercado.

2.4 Pregunta de investigación (Max 100 Palabras) - (*Primera entrega*)

¿Cómo predecir los precios de arriendo en las localidades de Bogotá utilizando un modelo de aprendizaje automático que tome como variables de entrada las características de las propiedades y datos del mercado inmobiliario en Bogotá?

2.5 Hipotesis (Max 100 Palabras) - (*Primera entrega*)

Los precios de arriendo en Bogotá se pueden predecir eficazmente a través de un modelo de aprendizaje automático que utiliza características de las propiedades y datos del mercado..

3 Reflexiones sobre el origen de datos e información (Max 400 Palabras) - (*Primera entrega*)

El origen de los datos es un factor crucial en cualquier proyecto de análisis de datos o inteligencia empresarial, especialmente en el contexto del mercado inmobiliario, donde la precisión y la actualidad de la información pueden tener un impacto significativo en las conclusiones y predicciones.

La calidad de los datos originales es fundamental. Los datos deben ser recopilados de fuentes confiables y autorizadas, como en este caso, el sitio web de FincaRaíz, que es reconocido en el sector inmobiliario de Colombia.

3.1 ¿Cual es el origen de los datos e información ? (Max 100 Palabras) - (*Primera entrega*)

El origen de los datos se va a obtener de la pagina web www.fincaraiz.com.co la cual es una de las principales páginas del mercado inmobiliario en Colombia y proporciona información detallada sobre propiedades en arriendo, que incluye precios, ubicación, características del inmueble y otras variables relevantes para tu análisis y modelado predictivo.

3.2 ¿Cuales son las consideraciones legales o éticas del uso de la información? (Max 100 Palabras) - (*Primera entrega*)

En cuanto a consideraciones legales, en Colombia, no existe una normativa acerca del Webscrapping, encontrándose en un limbo jurídico en cuanto sus alcances y beneficios o perjuicios particulares o empresariales; Y debido a que los datos que se van a extraer son de carácter público, no existe algún riesgo de vulneración de la ley 1581 de 2012 de protección de datos personales. De igual forma, no existen grandes consideraciones éticas en principio, ya que no son datos sensibles. Se van a utilizar únicamente con el fin de conocer el comportamiento de los valores de arrendamiento de las viviendas en Bogotá.

3.3 ¿Cuales son los retos de la información y los datos que utilizara en Integración y Automatización de Datos para IA? (Max 100 Palabras) - (*Primera entrega*)

Convertir los datos que se extraen en indicadores estándar, que permitan comparar los precios de los diferentes inmuebles como por ejemplo el valor por metro cuadrado en determinada localidad, y de igual forma parametrizar otros valores agregados que podrían influir en el precio como por ejemplo si tiene parqueadero, el número de baños, la antigüedad, la ubicación, entre otras. Tener la capacidad de discriminar y eliminar valores atípicos que no corresponden con la realidad, ya que algunos anunciantes utilizan técnicas para aumentar la visibilidad del inmueble en las búsquedas

3.4 ¿Que espera de la utilización de Integración y Automatización de Datos para IA para su proyecto? (Max 100 Palabras) -
(Primera entrega)

A través de la automatización, se espera obtener información con la frecuencia requerida, para observar las variaciones que se presentan en los precios de arrendamiento, y con base en estos hallazgos, brindar una herramienta que permita tomar decisiones en cuanto a los sitios mas recomendados para arrendar de acuerdo al costo, O a su ubicación, entre otras.

4 Diseño de integración y Automatización de Datos para IA (Diagrama) *(Primera entrega)*

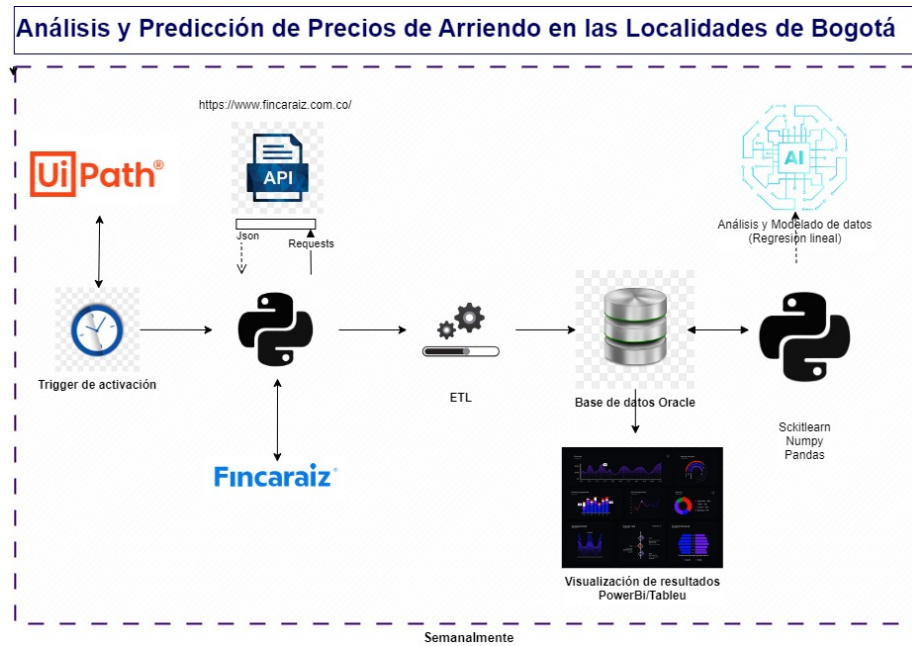


Figure 1: Diagrama flujo proyecto automatización

5 integración de Datos (*Segunda entrega*)

Vulputate nec hac convallis rutrum eu ante volutpat aliquam ullamcorper pulvinar tristique velit nulla, cubilia felis tempor aptent vitae rhoncus parturient euismod mauris potenti dignissim magna. Nunc nec cum quisque quam tincidunt mauris nascetur conubia placerat fusce consequat eget erat vulputate, est rhoncus etiam dignissim luctus volutpat facilisi molestie torquent at convallis accumsan. Magnis morbi auctor sapien fusce turpis leo ad libero vivamus, sem enim ultrices elementum curae accumsan vel hendrerit. Etiam elementum dui a sodales auctor lacus proin interdum, porttitor netus tortor blandit sociis facilisi ullamcorper, mi aenean euismod diam placerat dignissim class.

6 Automatización de Datos (*Segunda entrega*)

Vulputate nec hac convallis rutrum eu ante volutpat aliquam ullamcorper pulvinar tristique velit nulla, cubilia felis tempor aptent vitae rhoncus parturient euismod mauris potenti dignissim magna. Nunc nec cum quisque quam tincidunt mauris nascetur conubia placerat fusce consequat eget erat vulputate, est rhoncus etiam dignissim luctus volutpat facilisi molestie torquent at convallis accumsan. Magnis morbi auctor sapien fusce turpis leo ad libero vivamus, sem enim ultrices elementum curae accumsan vel hendrerit. Etiam elementum dui a sodales auctor lacus proin interdum, porttitor netus tortor blandit sociis facilisi ullamcorper, mi aenean euismod diam placerat dignissim class.

7 IA (*Segunda entrega*)

Vulputate nec hac convallis rutrum eu ante volutpat aliquam ullamcorper pulvinar tristique velit nulla, cubilia felis tempor aptent vitae rhoncus parturient euismod mauris potenti dignissim magna. Nunc nec cum quisque quam tincidunt mauris nascetur conubia placerat fusce consequat eget erat vulputate, est rhoncus etiam dignissim luctus volutpat facilisi molestie torquent at convallis accumsan. Magnis morbi auctor sapien fusce turpis leo ad libero vivamus, sem enim ultrices elementum curae accumsan vel hendrerit. Etiam elementum dui a sodales auctor lacus proin interdum, porttitor netus tortor blandit sociis facilisi ullamcorper, mi aenean euismod diam placerat dignissim class.

8 Proximos pasos (*Tercera entrega*)

9 Lecciones aprendidas (*Tercera entrega*)

10 Bibliografía