

## PROYECTO 2

### **Análisis Avanzado y Modelado Predictivo de Precios de Viviendas en Barcelona utilizando KNIME, AutoML y Power BI**

**(Fecha límite de entrega: 3 de marzo del 2024 hasta las 23:59)**

#### **Objetivo:**

Ampliar el análisis y modelado predictivo de precios de viviendas en Barcelona utilizando herramientas avanzadas como KNIME para el ETL y el análisis, Power BI para la visualización interactiva, y herramientas de AutoML como plataforma de Machine Learning Low-Code o No-Code. El objetivo es mejorar la precisión del modelo predictivo y proporcionar visualizaciones interactivas que faciliten la toma de decisiones.

#### **Requisitos:**

1. **KNIME:** Utilizar KNIME para realizar el proceso ETL (Extracción, Transformación y Carga) de los datos. La herramienta debe emplearse para integrar, limpiar y transformar los datos antes del análisis y modelado.
2. **Power BI:** Crear dashboards interactivos en Power BI para visualizar los resultados del análisis de datos y del modelo predictivo. Los dashboards deben incluir gráficos, mapas y filtros que permitan explorar los datos de precios de viviendas de manera intuitiva.
3. **AutoML:** Escoger una herramienta basada en los principios de AutoML (No-Code o Low-Code) como: BigML, KNIME, IBM Watson, Google AutoML, Azure ML, etc.

#### **Pasos del Proyecto:**

1. **Extracción y preparación de datos con KNIME:**
  - **ETL con KNIME:** Utilizar KNIME para extraer los datos de precios de viviendas desde el portal Fotocasa, transformarlos y cargarlos en un formato adecuado para el análisis. Esto incluye la limpieza

de datos, la integración de variables relevantes y la creación de un Data Lake para almacenar los datos procesados.

## **2. Análisis Exploratorio de Datos (EDA) con KNIME:**

- Realizar un análisis exploratorio de los datos utilizando KNIME. Esto incluye la creación de visualizaciones preliminares, identificación de patrones y detección de valores atípicos. KNIME proporcionará herramientas para explorar y entender mejor los datos antes de proceder al modelado.

## **3. Ética de la IA y Gobernanza de Datos:** El proyecto debe considerar la privacidad de los datos y el impacto ético de las predicciones. Se requiere que el equipo analice la gobernanza de los datos utilizados, asegurando que el tratamiento de la información cumple con las normativas vigentes (e.g., GDPR).

## **4. Desarrollo del modelo predictivo con herramientas AutoML:** Emplear a elegir por el equipo una herramienta de Auto Machine Learning que permita poder entrenar modelos de tipo regresor, analizar sus métricas y, cuando se disponga de resultados aceptables, exportar dichas predicciones para emplearlas a posteriori en el informe de Power BI

## **5. Visualización de datos con Power BI:**

- **Creación de dashboards en Power BI:** Desarrollar dashboards interactivos en Power BI para presentar los resultados del análisis de datos y las predicciones del modelo (Predicciones exportadas desde la herramienta de AutoML). Incluir visualizaciones como gráficos de distribución, mapas de calor por ubicación y filtros interactivos para explorar diferentes aspectos de los datos.

## **6. Entrega y documentación:**

- **Integración en GitHub:** Documentar todo el trabajo en un repositorio de GitHub, incluyendo el flujo de trabajo de KNIME, archivos de configuración (cuidado con exponer información privada, recordad que toda información que se desee no exponer se puede incluir en el archivo .gitignore), modelos exportados y los dashboards de Power BI (también cabe la posibilidad de exportar las métricas DAX) así como la documentación del proceso.
- **Informe detallado:** Redactar un informe que explique el proceso de ETL, el análisis exploratorio, la comparativa de modelos y, la selección del modelo predictivo y la visualización de los resultados. Incluir capturas de pantalla de los dashboards en Power BI y una descripción detallada de las conclusiones obtenidas.

**Resultados Esperados:**

1. **ETL y EDA de los datos:** Un proceso de Extracción, Transformación y Carga sobre los datos, obteniendo las potenciales tablas dimensionales y tablas de hechos, así como todo el proceso de Análisis Exploratorio de los Datos.
2. **Modelo Predictivo Mejorado:** Un modelo predictivo refinado utilizando herramientas de AutoML con métricas de rendimiento claramente documentadas.
3. **Dashboards Interactivos en Power BI:** Dashboards completos que permitan una visualización interactiva de los datos y resultados del modelo.

**Código y Documentación en GitHub:** Repositorio bien organizado con todo el código, la documentación del proyecto (README.md) y los informes necesarios.

**Dataset:**

- **Barcelona\_Fotocasa\_HousingPrices\_Augmented.csv**