Proyecto de desarrollo de sistema de IA basado en alquileres de pisos – Primera fase

1. **Explicación del proyecto**

En la primera fase de este proyecto vamos a elaborar un sistema de inteligencia artificial apoyado en un dataset que incluye datos acerca de alquileres en la capital de España, Madrid. Este documento contiene datos importantes acerca del mercado de venta de pisos, lo que nos llevará a desarrollar un sistema de inteligencia artificial que nos facilite ciertos procesos que, de otra forma, no serían nada sencillos.

En nuestro caso, el sistema va a analizar y anticipar posibles tendencias que se vayan a producir en el mercado de alquileres de Madrid.

El sistema será capaz de analizar similitudes, pisos disponibles y demanda de los mismos. Ayudará a inversores a saber en qué barrios puede ser más rentable invertir y a los clientes a conocer zonas donde encontrar alquileres asequibles.

Cumpliremos las normativas legales en todo momento y seguiremos los principios de ética.

1. **Valoración de que no es una opción de “riesgo inaceptable”**

Como ya hicimos en la actividad 2, nos apoyaremos en los artículos 5, 6, 8 y 9 para realizar la validación.

El artículo nº 5 menciona prácticas prohibidas, que no se pueden realizar a la hora de implantar un sistema de IA. En nuestro caso, el sistema tratará datos introducidos en un dataset que han sido publicados de manera anterior en una página. Esto nos lleva a asegurar que en ningún momento vulneraremos ninguna de las mencionadas en este artículo.

En el anexo III del artículo nº 6 se califican los sistemas en función de si son de riesgo aceptable o no. Comprobamos que en ninguno de los apartados en los que califican los sistemas aparecería el nuestro.

Siguiendo lo que dice el artículo nº8, garantizaremos el cumplimiento de los sistemas de IA de alto riesgo a que se refiere el apartado 1 introduciremos “los procesos de prueba y presentación de información necesarios, y la información y la documentación que faciliten con respecto a su producto en documentación y procedimientos que ya existan y exijan los actos legislativos de armonización de la Unión enumerados en el anexo I”.

Y finalmente, el artículo nº 9, que pide que gestionemos los riesgos utilizando un sistema que durará todo el ciclo de vida del sistema de alto riesgo evaluando y realizando actividades periódicas. Buscaremos una persona responsable, teniendo en cuenta sus conocimientos técnicos, experiencia, educación y formación, con vistas a eliminar o reducir los riesgos asociados a la utilización del sistema de IA de alto riesgo.

Como es de esperar aseguraremos la protección de datos siguiendo al RGPD y manteniendo el anonimato de estos.

En cuanto a posibles riesgos en el sesgo de nuestro sistema, identificaremos en todo momento posibles análisis que sean erróneos o que den información poco relevante o falsa acerca de los alquileres. Los posibles riesgos legales o éticos estarán permanentemente revisados en auditorias.

Por último, garantizaremos la transparencia de nuestro sistema haciendo que sea perfectamente analizable por cualquier usuario. En todo momento mantendremos una supervisión de cualquier análisis generado por el mismo.

1. **Objetivos del negocio conforme a la primera fase de CRISP DM**

Nuestro sistema va a ayudar a analizar las posibles tendencias del mercado inmobiliario en Madrid, facilitando así procesos de inversión y alquiler. Nuestra IA servirá para analizar patrones dentro del dataset y devolvernos tendencias, siguiendo algunas de las columnas más importantes del dataset utilizado.

La IA predecirá posibles zonas con un aumento de ventas, posibilidad de futuros alquileres o venta de inmuebles. Nos apoyaremos del dataset y de herramientas que usan técnicas de auto regresión como ARMA, ARIMA, SARIMA, TBATS en nuestro caso usaremos ARIMA.

Los criterios de éxito los dividiremos en:

Cualitativos

* Generar tendencias de mercado que ayuden a los inversores a decidir el mercado correcto.
* Reducción de los alquileres no rentables o que ya hayan sido comprados por alguien.
* Previsión de posibles alquileres para clientes que busquen algo específico.

Cuantitativos

* Aumento de los beneficios de los inversores en un 50%
* Identificación mayor del 97% de alquileres no útiles o viables.
* Generar posibles alquileres rentables en segundos.

1. **Evaluación de la situación conforme a la segunda fase de CRISP DM**

El estudio del arte en este caso nos lleva a buscar en internet herramientas relacionadas con el mercado de alquileres de inmuebles. Algunas de las herramientas que he encontrado son las siguientes: Kensho, PriceLabs y Inmovilla.

Los datos disponibles en este caso serían principalmente obtenidos del dataset. Igualmente, para la mejora de este sistema se podría implementar una herramienta que, cada un determinado periodo de tiempo, obtenga datos de las principales páginas inmobiliarias y los importe dentro de un dataset. Para valorar el coste-beneficio de nuestro proyecto nos serviremos de una tabla que represente estos valores, siempre teniendo en cuenta los aspectos éticos y legales que cumplan la protección de datos recogida acorde a la RGPD.

|  |  |  |
| --- | --- | --- |
| ITEM | COSTO (€) | BENEFECIO (€) |
| Fase de Diseño e Investigación | | |
| Investigación de mercado | 2000 | 15000 |
| Diseño de la interfaz de usuario | 2150 | 10000 |
| Diseño de la experiencia de usuario | 3000 | 17500 |
| Fase de Desarrollo | | |
| Desarrollo de IA | 20000 | 130000 |
| Desarrollo de la herramienta | 14000 | 65000 |
| Pruebas y correcciones | 4000 | 9000 |
| Fase de Lanzamiento | | |
| Marketing y publicidad | 2500 | 5500 |
| Soporte y servicio al cliente | 1.400 | 3000 |
| Costos recurrentes | | |
| Mantenimiento y actualizaciones | 750/año | 4000/año |
| Servicios en la nube | 1.200/año | 15000/año |
| Gastos legales y éticos |  |  |
| Consultoría Legal y Ética | 2500 | 0 |
| Auditoría de Conformidad Legal | 3500 | 0 |
| Cumplimiento GDPR | 2000 | 0 |
| Revisión y Supervisión Humana | 2500 | 0 |
| Totales | 37500 | 225000 |

En este caso hemos aumentado el gasto en la fase de desarrollo, incorporando una persona más en cada puesto científico de datos e ingeniero de software. Al final necesitaremos gente capaz de utilizar nuestra herramienta de regresión y que puedan llevar a cabo el modelo del sistema de IA. Ampliando la plantilla, inicialmente conseguiremos un sistema mucho mejor en menos tiempo. Los costes en gastos legales y éticos los mantendremos, ya que es importante cumplir con las leyes de protección y distintas normativas, además de mantener bajo supervisión en todo momento los resultados.

1. **Explicación del modelado, análisis del dataset y pequeño plan de proyecto**

En nuestro proyecto nos ayudaremos de ARIMA. Esta herramienta es clave para predecir tendencias de mercado. Nos podrá ayudar a predecir precios y tendencias gracias al manejo de valores históricos. También servirá para poder brindar a inversores información acerca de cuándo es mejor invertir, reducir o aumentar precios en función de la época del año o valorar qué está haciendo que cambie la demanda en los alquileres.

El dataset que estamos utilizando está bastante obsoleto, ya que la mayoría de datos que se encuentran en él son de pisos que en la actualidad ya han sido vendidos. Igualmente, la estructura que presenta nos servirá de ayuda, pues tiene una gran cantidad de columnas con datos importantes para nuestro sistema.

En nuestro caso utilizaremos las columnas de: price, floor\_area,type,district y subdistrict. Estas nos ayudarán a generar unos datos concretos. Finalmente, eliminaríamos la columna de garden\_area ya que no muestra datos.

Respecto al plan de proyecto, apoyado en nuestra tabla de costes/beneficios: Los primeros meses haremos un trabajo con el dataset, analizándolo, realizando la limpieza y aportando nuevos datos. Los siguientes meses serán la fase de desarrollo, en esta fase ajustaremos ciertas partes del modelo, nos apoyaremos en la herramienta ARIMA y dejaremos una interfaz agradable de cara al usuario.

Seguiremos con una fase de corrección de errores, pruebas y validación de aspectos legales y éticos. Por último, realizaremos la fase de lanzamiento a ciertos usuarios para poder recopilar información y mejorar su uso en posteriores actualizaciones.