

Título de la solución

Solución de Inteligencia Artificial y Big Data para facilitar las tareas de reconocimiento y validación de documentos en los canales digitales

Descripción de la idea

Para responder al reto planteado que consiste en facilitar las tareas de reconocimiento y validación de documentos en los canales de comunicación digitales y mejorar así el servicio de atención a los ciudadanos del servicio municipal de gestión del ciclo integral del agua, proponemos el desarrollo de una solución software conversacional basada en las nuevas capacidades que ofrece la Inteligencia Artificial generativa en el ámbito del procesamiento del lenguaje natural (NLP).

El objetivo es automatizar en la medida de lo posible tareas rutinarias y repetitivas relacionadas con el reconocimiento y validación de los documentos aportados por los clientes en la contratación o cambio de titularidad de los suministros de agua. Esto es, por una parte, validar documentos de tipo identificación (DNI, CIF), documento que acredita uso de la vivienda o local (escritura, contrato de alquiler), certificado de cuenta bancaria para la domiciliación de los recibos u otros documentos relevantes, y por otra parte, extraer los datos clave de esos documentos para poder comprobar su coherencia así como rellenar los datos de registro necesarios para realizar el procedimiento administrativo (cambio titularidad, alta, baja, etc.).

Además, el sistema debe ser capaz de incorporar nuevos tipos de documentación para adaptarse a los reglamentos y ordenanzas que regulan la prestación del servicio de ciclo integral del agua en diferentes municipios.

Este tipo de proyectos, habitualmente, se aborda en tres fases que se centran en el **diagnóstico** (análisis), **habilitación** (diseño) y **ejecución** (implementación) donde, partiendo de la evaluación y estudio de la situación inicial, se pretende diseñar y poner en marcha la arquitectura de solución considerada óptima en base a los requisitos específicos definidos y recogidos durante la fase de diagnóstico:

1. Fase de diagnóstico. Se abordarán las siguientes tareas:
 - a. Análisis de la situación actual y de las necesidades a nivel de orígenes de datos (fuentes, protocolos de comunicación, tipo de datos, periodicidad o frecuencia de muestreo, volumetría, etc.), almacenamiento y explotación de la información. Se estudiarán los distintos subsistemas a integrar, así como las necesidades específicas de cada uno de ellos con respecto a los datos que ofrecen y su interoperabilidad.
 - b. Análisis de los requisitos del sistema de soporte a la toma de decisiones a implementar, desde una perspectiva de explotación de los datos y posibles usuarios potenciales de dicho sistema.
 - c. Elaboración de un informe de la situación actual, requisitos detectados y diagnóstico de futuros pasos.

2. Fase de habilitación. El objetivo de esta fase es realizar el diseño de arquitectura de los componentes que integrarán la solución final, en base a los requisitos capturados durante la fase anterior. Se abordarán las siguientes tareas:
 - a. Análisis de los componentes y servicios necesarios para la consecución de los objetivos del proyecto en base a un presupuesto previamente establecido y según lo estipulado en la fase de diagnóstico.
 - b. Diseño de la arquitectura de la solución final.
 - c. Definición de los cuadros de mando a desarrollar con los KPIs e indicadores detectados de interés
 - d. Estudio de viabilidad para la aplicación de modelos de IA/ML en base a los datos recogidos y del histórico disponible.
3. Fase de ejecución. En base al alcance y los objetivos establecidos, se llevará a cabo la implementación de los componentes, servicios, así como las comunicaciones a establecer para la puesta en marcha del sistema de apoyo a la gestión documental. Las tareas a realizar consistirán en:
 - a. La implementación de las ETLs necesarias para conectar con los orígenes de datos, recuperar la información considerada relevante y almacenar dichos datos en el espacio de almacenamiento previamente definido.
 - b. Creación del espacio de almacenamiento, así como de los mecanismos necesarios para acceder al mismo, en base a las necesidades de explotación previamente definidas.
 - c. Creación de los cuadros de mando definidos para la:
 - i. Visualización de la información en tiempo real.
 - ii. Análisis en base a histórico (indicadores y KPIs considerados de interés).
 - d. Si durante la habilitación se ha considerado viable el uso de técnicas de IA/ML para la elaboración de un modelo de este tipo, se llevará a cabo el análisis exploratorio de datos y el entrenamiento de dichos modelos de IA/ML.

Los perfiles técnicos que participarán en el proyecto serán:

1. **Arquitecto Big Data:** diseña arquitecturas que permiten la gestión de grandes volúmenes de datos que no pueden ser tratados de manera convencional. Se encarga de definir la estrategia que se aplicará con respecto al gobierno y seguridad de los datos.
2. **Ingeniero Big Data:** participa en las etapas donde se involucran los datos, siendo éstos el centro de su actividad. Se enfoca en el diseño, desarrollo y mantenimiento de los procesos necesarios para la extracción, carga, almacenamiento y procesamiento de los datos.
3. **Científico de Datos:** especialista en análisis avanzado de datos mediante técnicas de Machine Learning y tecnologías Big Data.
4. **Desarrollador software fullstack:** especialista en el desarrollo de aplicaciones que integran la implementación del frontend con la programación del backend de soporte a los procesos.

Hitos y tiempos hasta implementación

El inicio/fin de los trabajos vendrá condicionado por las necesidades que el cliente plantee, la disponibilidad de la información necesaria para poder llevar a cabo las diferentes tareas planteadas, así como la disponibilidad del equipo asignado por parte del cliente para la participación en las tareas de diagnóstico y análisis. No obstante, a modo orientativo, se plantea el siguiente cronograma de 8 meses de duración (M1-M8):

- Diagnóstico (M1-M3)
- Habilitación (M3-M5)
- Ejecución (M5-M8)

	Meses							
	1	2	3	4	5	6	7	8
Diagnóstico								
Habilitación								
Ejecución								