**PRELIMINAR PLAN DE TRABAJO PROYECTO**

Proyecto: Montar un modelo conceptual para potencializar la implementación de herramientas de análisis de texto para la detección y clasificación automática de términos sensibles en documentos y metadatos de la organización.

Objetivo del proyecto: Potencializar la implementación de herramientas de análisis de texto para la detección y clasificación automática de términos sensibles en documentos y metadatos de la organización, con el objetivo de mejorar la calidad de los datos y asegurar la protección de la información sensible.

Nicho inicial: ADRES, Colmena; Contraloría general de la Nación

Recursos: Se usará un repositorio de Azure en GIT, en el cual se hace análisis de riesgo crediticio usando AML.

1. Investigación y análisis de requerimientos (40 horas):

* Revisión de los documentos y metadatos de la organización para determinar qué términos sensibles se deben detectar y clasificar automáticamente.
* Identificación de los recursos de Azure necesarios para el proyecto.
* Definición de los requisitos y objetivos del proyecto.
* Adaptación inicial del proyecto para usar Azure Purview.

1. Configuración de los recursos de Azure (20 horas):

* Configuración de Microsoft Purview para el descubrimiento de datos y metadatos.
* Configuración de Azure Synapse Analytics para el procesamiento y análisis de los datos y metadatos descubiertos por Microsoft Purview.
* Configuración de Azure Machine Learning para la construcción y entrenamiento de modelos de aprendizaje automático.
* Configuración de Azure OpenAI para la integración de tecnologías avanzadas de procesamiento de lenguaje natural.

1. Preparación de los datos (30 horas):

* Limpieza y preprocesamiento de los datos no estructurados y metadatos para su procesamiento en Azure Synapse Analytics y Azure Machine Learning.
* Conversión de los datos en un formato compatible con los recursos de Azure.

1. Construcción y entrenamiento de modelos de aprendizaje automático (60 horas):

* Selección de los algoritmos de aprendizaje automático más apropiados para el proyecto.
* Construcción y entrenamiento de los modelos de aprendizaje automático utilizando los recursos de Azure Machine Learning y Azure OpenAI.

1. Validación y ajuste del modelo (40 horas):

* Validación de los modelos de aprendizaje automático construidos.
* Ajuste de los modelos de aprendizaje automático para mejorar su precisión y rendimiento.

1. Despliegue del modelo (20 horas):

* Despliegue del modelo de aprendizaje automático en la infraestructura de Azure.

1. Pruebas y ajustes finales (30 horas):

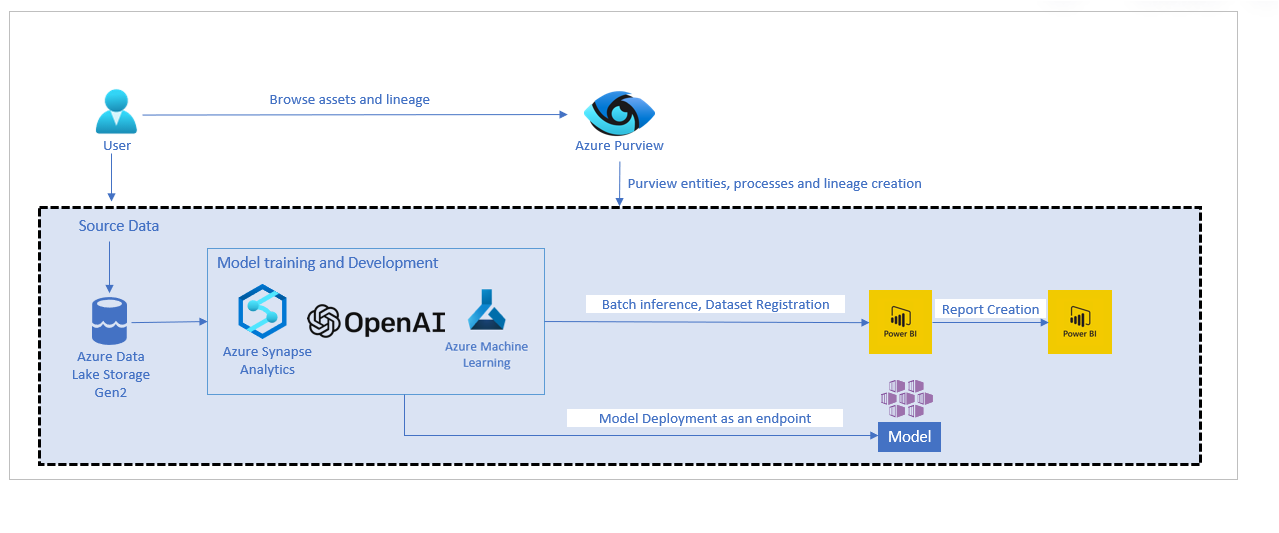
* Pruebas finales del modelo de aprendizaje automático desplegado.
* Ajustes finales del modelo de aprendizaje automático.

**CASO DE USO HIPOTETICO**

Supongamos que la empresa XYZ quiere mejorar la calidad de sus datos y asegurar la protección de su información sensible mediante la implementación de herramientas de análisis de texto para la detección y clasificación automática de términos sensibles en sus documentos y metadatos. El siguiente es un caso de uso detallado para montar un proyecto que cumpla con estos objetivos utilizando los recursos de Azure mencionados:

1. Identificación de los documentos y metadatos relevantes: Se debe identificar cuáles son los documentos y metadatos que contienen información sensible y que necesitan ser analizados. Se debe definir una estrategia para identificarlos y recopilarlos, incluyendo la implementación de herramientas de clasificación y etiquetado.
2. Integración de los datos en Microsoft Purview: Una vez que se han identificado los documentos y metadatos relevantes, se deben integrar en Microsoft Purview, una plataforma de gobierno de datos de Azure. Esto permitirá a la empresa gestionar y catalogar los datos de manera eficiente, y asegurará que los datos sean precisos y estén disponibles para su análisis.
3. Preparación de los datos en Azure Synapse Analytics: Los datos integrados en Microsoft Purview deben prepararse para su análisis en Azure Synapse Analytics, una plataforma de análisis de datos de Azure. Esto incluye la limpieza de los datos, la normalización y la creación de conjuntos de datos de entrenamiento y validación.
4. Implementación de modelos de aprendizaje automático en Azure Machine Learning: Se deben desarrollar modelos de aprendizaje automático para la detección y clasificación de términos sensibles en los documentos y metadatos. Esto se puede hacer utilizando Azure Machine Learning, que ofrece herramientas y servicios para desarrollar y entrenar modelos de aprendizaje automático.
5. Implementación de Azure OpenAI: La empresa puede aprovechar Azure OpenAI para mejorar la precisión de los modelos de aprendizaje automático mediante la generación de lenguaje natural. Esta herramienta utiliza tecnologías avanzadas de procesamiento del lenguaje natural para mejorar la comprensión y generación de texto.
6. Despliegue y puesta en producción del modelo conceptual: Una vez que se han desarrollado y entrenado los modelos de aprendizaje automático, se deben desplegar y poner en producción en la plataforma de análisis de datos Azure Synapse Analytics. Se deben monitorear los resultados y realizar ajustes en el modelo para mejorar su precisión y eficacia.
7. Mantenimiento y actualización continua del modelo: Para garantizar que el modelo siga siendo preciso y eficaz, es importante realizar un mantenimiento y actualización continuos. Esto incluye la adición de nuevos términos sensibles y la actualización de los modelos para adaptarse a los cambios en los datos y el entorno.

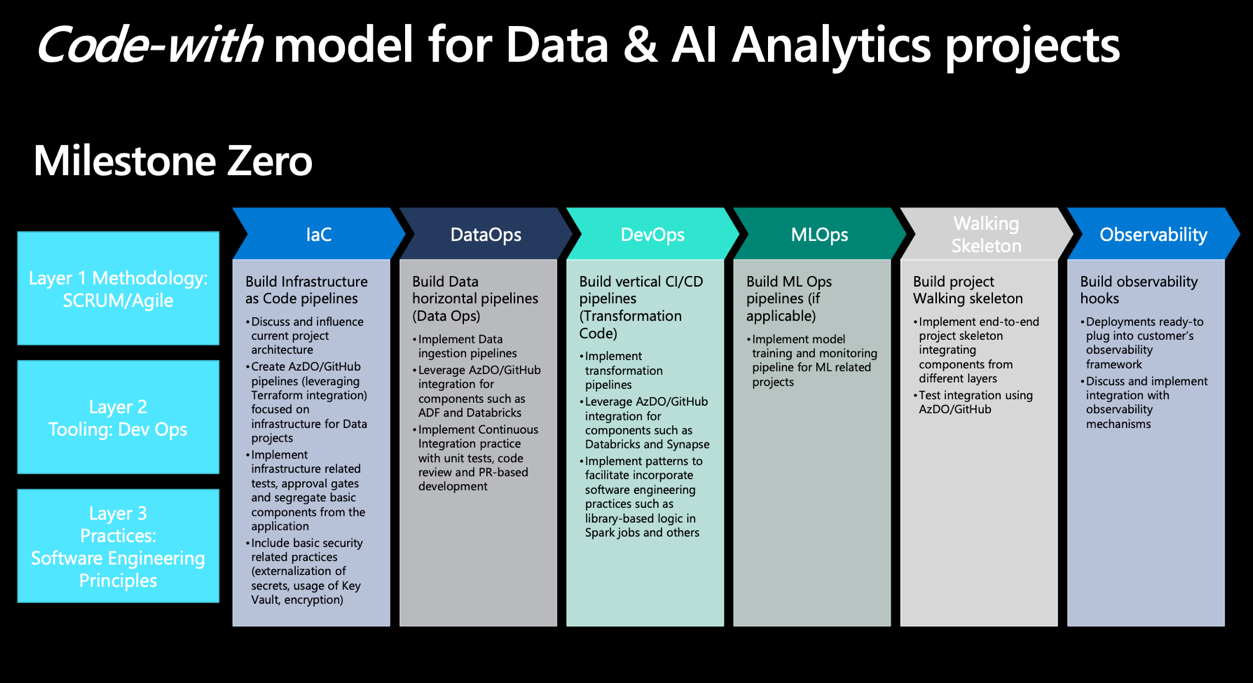
**ARQUITECTURA**



**LINAJE DEL PROCESO PURVIEW ML**



**El Laboratorio se desarrollara bajo la metodología Agile/SCRUM Dev Squad.**



**PLAN DE TRABAJO**

****