Imagen que contiene Logotipo

Descripción generada automáticamente

**OPERACIONES DE APRENDIZAJE AUTOMÁTICO**

**Documento de Gobernanza**

**Profesor:** Dr. Gerardo Rodríguez Hernández

**Profesor Tutor:** Mtro. Francisco Javier López Tiro

**Equipo #01** | Integrantes del Equipo:

Daniel Acevedo Sainos - **A01795496**

Luis Alejandro Aguilar Diaz - **A01795362**

Marcos Eduardo García Ortiz - **A01276213**

Héctor Raúl Peraza Alavez - **A01795125**

Juan Manuel Rodríguez Mateos - **A01794890**

Andrea Monserrat Ruiz Gómez **- A01794631**

03 de noviembre de 2024

**Documento de Gobernanza del Modelo de Predicción de Diabetes**

**1. Introducción**

Este documento describe las políticas y procedimientos de gobernanza para nuestro modelo de Machine Learning (ML) de predicción de diabetes. El objetivo principal de este modelo es predecir con precisión la diabetes en pacientes, con el fin de ayudar en el diagnóstico temprano y la evaluación de riesgos. La gobernanza del modelo asegura su calidad, seguridad, confiabilidad y cumplimiento, fomentando su uso responsable y ético.

**2. Principios de Gobernanza**

Los siguientes principios guían la gobernanza de este modelo de ML:

* Precisión: El modelo debe ser preciso y confiable en sus predicciones.
* Equidad: El modelo debe ser justo y evitar sesgos que puedan discriminar a ciertos grupos demográficos.
* Transparencia: El proceso de desarrollo y las decisiones del modelo deben ser transparentes y comprensibles.
* Privacidad: Los datos de los pacientes deben ser protegidos y utilizados de manera responsable.
* Responsabilidad: Se debe rendir cuentas del desarrollo, implementación y uso del modelo.

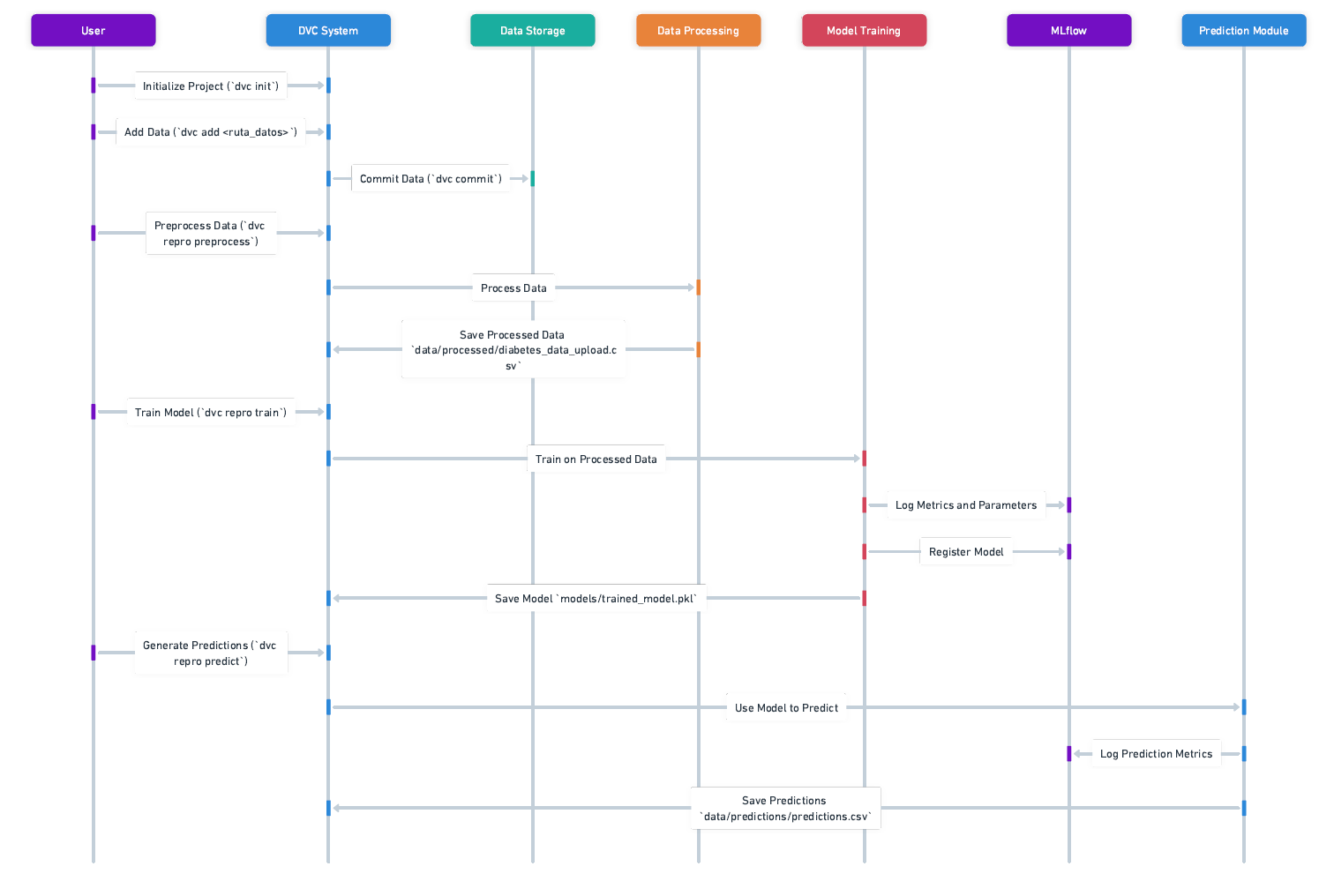
**3. Políticas y Procedimientos**

**3.1 Gestión del Modelo (Model Governance)**

* Versionamiento: Se utilizará DVC para el versionamiento del modelo, incluyendo el código fuente, los datos de entrenamiento, los hiperparámetros y las métricas de rendimiento. Cada versión del modelo será etiquetada y almacenada de forma segura.
* Monitoreo del rendimiento: Se implementará un sistema de monitoreo continuo para rastrear el rendimiento del modelo en producción. Se definirán métricas clave (precisión, F1-score, AUC) y umbrales de alerta para detectar la degradación del modelo.
* Reentrenamiento: El modelo se reentrenará periódicamente con nuevos datos para mantener su precisión y adaptarse a posibles cambios en los patrones de la enfermedad.
* Evaluación de sesgos: Se analizarán los datos de entrenamiento y las predicciones del modelo para identificar posibles sesgos. Se utilizarán técnicas de mitigación de sesgos si se detectan.

**3.2 MLOps**

* MLflow: Se utilizará MLflow para el seguimiento de experimentos, el registro de parámetros, métricas, artefactos y la gestión de modelos.
* DVC: DVC se utilizará para la gestión de datos, el versionamiento de código y la gestión de pipelines de ML.
* CI/CD: Se implementará un pipeline de CI/CD para automatizar el entrenamiento, las pruebas y el despliegue del modelo.
* Diabetes MLOps Pipeline with DVC and MLflow:



**3.3 Estándares de Código**

* Se definirán estándares de codificación para garantizar la legibilidad, mantenibilidad y calidad del código.
* Se utilizarán herramientas de análisis de código estático (linters) para hacer cumplir los estándares de código.
* (Referido al repositorio: https://github.com/LAguilar35/diabetes\_mlops/blob/main/docs/docs/CODE\_STANDARDS.md).

**3.4 Verificaciones Éticas**

* Se realizará una evaluación ética del modelo para identificar posibles impactos negativos.
* Se documentarán las consideraciones éticas y se tomarán medidas para mitigar cualquier riesgo.
* (Referido al repositorio: https://github.com/LAguilar35/diabetes\_mlops/blob/main/docs/docs/ETHICAL\_GUIDELINES.md).

**3.5 Documentación**

* Se documentará todo el proceso de desarrollo del modelo, incluyendo la metodología, los datos, las decisiones de diseño y los resultados de las evaluaciones.
* Se mantendrá un registro de cambios para documentar las actualizaciones y modificaciones del modelo.

**3.6 Control de Acceso**

* Se implementarán controles de acceso para garantizar que solo las personas autorizadas puedan acceder al código, los datos y el modelo.

**3.7 Gestión de Riesgos**

Plan de riesgos: Se elabora un plan de riesgos que identifique, evalúe y mitigue los posibles riesgos asociados al modelo de ML. Este plan incluye:

* Identificación de riesgos: Se identificarán los posibles riesgos, incluyendo:
* Riesgos de precisión: Errores en las predicciones del modelo que puedan tener consecuencias negativas.
* Riesgos de predicción: Generación de predicciones incorrectas.
* Riesgos regulatorios: Cumplimiento con regulaciones.
* Riesgos Operacionales.
* Evaluación de riesgos: Se evaluará la probabilidad de ocurrencia y el impacto de cada riesgo.
* Mitigación de riesgos: Se definirán estrategias para mitigar cada riesgo identificado.
* (Referido al repositorio: https://github.com/LAguilar35/diabetes\_mlops/blob/main/docs/docs/RISK\_MANAGEMENT\_PLAN.md).

**4. Roles y Responsabilidades**

* Científico de datos: Responsable del desarrollo, entrenamiento y evaluación del modelo.
* Ingeniero de MLOps: Responsable de la implementación, el despliegue y el monitoreo del modelo.
* Equipo de ética: Responsable de la evaluación ética del modelo y la mitigación de riesgos.

**5. Cumplimiento**

Este modelo se desarrollará y utilizará de acuerdo con las leyes y regulaciones aplicables, incluyendo las leyes de protección de datos y las regulaciones de dispositivos médicos.

**6. Revisión y Actualización**

Este documento de gobernanza se revisará y actualizará periódicamente para reflejar los cambios en el modelo, las mejores prácticas y las regulaciones.