

OPERACIONES DE APRENDIZAJE AUTOMÁTICO

FASE 1

Avance de Proyecto

Equipo 6:

Andrea Cantú Martínez A01235000

Oscar Becerra Alegria A01795611

Jorge Reyes Londono A01794421

Henry Aranzales Lopez A01794020

Fernando Torres Faúndez A01796759

CONTENIDO



Da clic sobre cada título para acceder al contenido



VIDEO DE EQUIPO



ACCESO A
REPOSITORIO GITHUB



INTRODUCCIÓN



DESCRIPCIÓN DEL
PROBLEMA



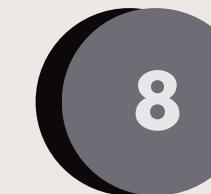
ASPECTOS CLAVE



ROLES



METODOLOGIA



PLAN DE ACCIÓN



RESPONSABILIDADES



FASES DE LA
CONSTRUCCIÓN



CONTROL DE VERSIONES



CONCLUSIONES

INTRODUCCIÓN

- Este proyecto tiene como objetivo desarrollar un modelo de machine learning que prediga la supervisión de pacientes pediátricos tras un trasplante de médula ósea, utilizando el dataset de la UCI.
- Aplicando una arquitectura de MLOps, el proyecto busca automatizar y escalar el ciclo de vida del modelo, garantizando su implementación continua y actualizada en producción.
- El modelo proporcionará una herramienta predictiva que apoyará a los profesionales médicos en la toma de decisiones basadas en datos clínicos, con el fin de optimizar los resultados de los trasplantes.



PROBLEMA

Predicción del Tiempo de Supervivencia en Trasplantes de Médula Ósea

- El desafío principal de este proyecto es predecir la supervivencia de los pacientes pediátricos que han recibido un trasplante de médula ósea.
- Resolver este problema permitirá a los médicos ajustar los tratamientos post-trasplante y mejorar las tasas de supervivencia de los pacientes, basándose en predicciones personalizadas.

ASPECTOS CLAVE

01

Dataset

Incluye datos de 187 pacientes pediátricos que recibieron un trasplante de células madre, con información clínica clave para el análisis del tiempo de supervivencia post-trasplante.

02

Hipótesis

Un mayor número de células CD34+ por kilogramo de peso del paciente está relacionado con un mayor tiempo de supervivencia sin eventos adversos significativos.

03

Pacientes

Los pacientes incluyen niños con diversas enfermedades, como leucemia, anemia aplásica, linfomas y otras condiciones que requieren trasplantes de médula ósea como tratamiento.

ROLES MLOps



Oscar
Subject Matter
Experts

Risk Manager



Andrea
Data Scientist



Jorge
ML Architect



Fernando
Data Engineer



Henry
DevOps

CRISP-DM

METODOLOGÍA

- Comprensión del negocio
- Comprensión de los datos
- Preparación de los datos
- Modelado
- Evaluación
- Despliegue
- Monitoreo y Mantenimiento

PLAN DE ACCIÓN

Identificación de problemas de negocio

Exploración y preprocesamiento de datos

Desarrollo de modelos de machine learning

Diseño y escalabilidad de la infraestructura

Optimización e implementación del modelo en producción

Monitoreo de riesgos y cumplimiento normativo

RESPONSABLES

Identificación de problemas de negocio

Subject Matter Experts

Exploración y preprocessamiento de datos

Data Engineer

Desarrollo de modelos de machine learning

Data Scientist

Diseño y escalabilidad de la infraestructura

DevOps

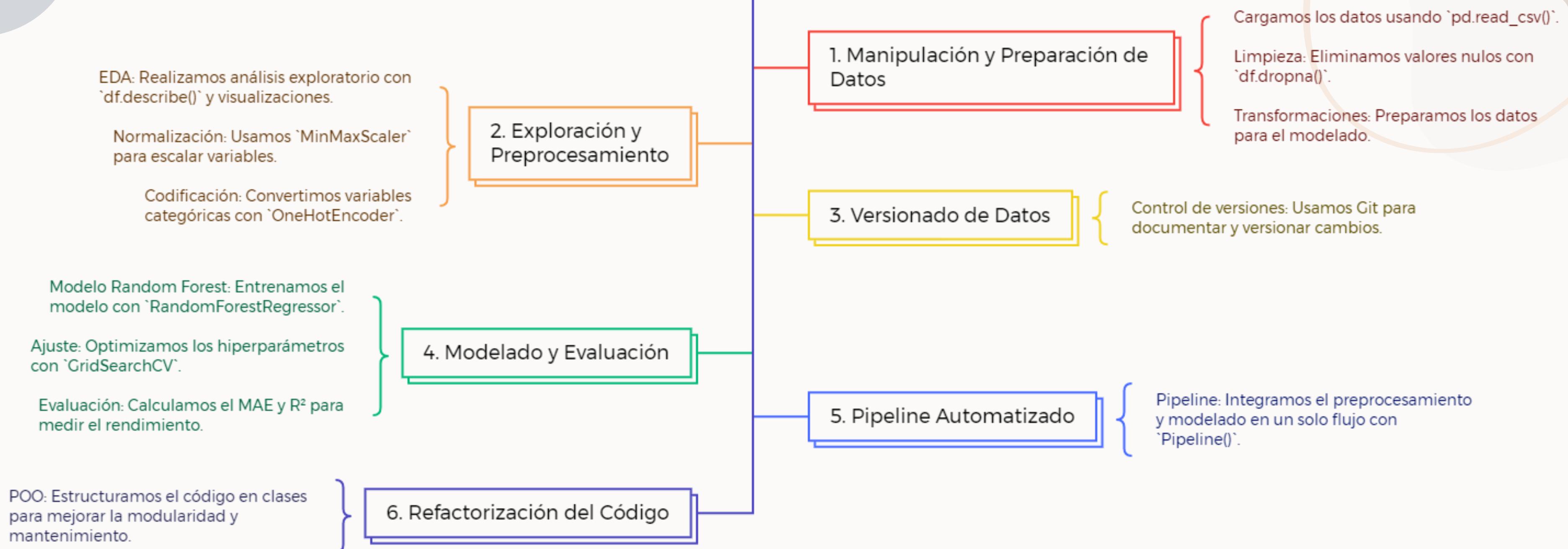
Optimización e implementación del modelo en producción

ML Architect

Monitoreo de riesgos y cumplimiento normativo

Risk Manager

Proceso de Machine Learning



USO DE GIT Y GITHUB

Git nos permitió versionar el código de manera incremental, lo que nos ayudó a realizar cambios y mejoras sin perder el historial de modificaciones. Cada miembro del equipo pudo trabajar en diferentes aspectos del proyecto sin conflictos, gracias al uso de ramas y la integración continua.

GitHub fue el repositorio central donde almacenamos nuestro código, lo que facilitó la colaboración y la revisión de cada cambio realizado. Con GitHub, podíamos hacer "pull requests" y recibir retroalimentación antes de fusionar los cambios a la rama principal, mejorando la calidad del código.

USO DE DVC

La integración de **DVC** fue clave para mantener la consistencia entre las versiones de los datos y el código. Cada versión del conjunto de datos fue asociada con una versión del código, permitiéndonos reproducir experimentos de manera precisa. Esto es especialmente importante en proyectos de Machine Learning, donde los modelos dependen directamente de los datos de entrada.

BENEFICIOS DEL TRABAJO EN EQUIPO

Colaboración: Al versionar tanto el código como los datos, todos los miembros del equipo pudieron trabajar en paralelo, sabiendo que cualquier cambio realizado estaba respaldado y registrado.

Trazabilidad: Con el historial de cambios en Git y la capacidad de versionar los datos con DVC, siempre supimos qué modificaciones se realizaron y cómo afectaban al proyecto. Esto facilitó la depuración y mejora continua del pipeline de ML.

Facilidad de reproducción: Gracias a la estructura de versionado, cualquier miembro del equipo podía volver a una versión específica del proyecto para reproducir resultados, lo que mejoró la transparencia y confiabilidad de los experimentos.

CONCLUSIONES

La calidad del
preprocesamiento es
esencial.

La evaluación y
ajuste de modelos
mejoró
significativamente el
rendimiento

La calidad del
preprocesamiento es
esencial.

REFLEXIÓN FINAL

El versionado fue esencial para mantener el orden y la eficiencia en nuestro proyecto de MLOps. Nos permitió trabajar de manera colaborativa y coordinada, evitando conflictos o la pérdida de trabajo, lo que resultó en un flujo de desarrollo más ágil. Cada miembro del equipo pudo enfocarse en sus responsabilidades, contribuyendo al proyecto sin interferir en el trabajo de los demás. Este enfoque también fomentó la comunicación y la retroalimentación continua, asegurando que cada cambio o mejora fuera revisada y validada, lo que contribuyó a un producto final más robusto y bien estructurado. En resumen, el versionado no solo facilitó la organización del trabajo, sino que también mejoró la sinergia y el aprendizaje colectivo dentro del equipo, impulsando la calidad y el éxito del proyecto.

MUCHAS
GRACIAS

