



CIENCIA DE DATOS E IA

Cohorte 2022

Aprendizaje automático

Minería de Datos

Práctica Profesionalizante

Segunda evidencia

Profesor Facundo Cúneo

Integrantes Grupo 4

- 1.- Sol Figueroa
- 2.- Walter Nieto
- 3.- Eduardo Figueroa
- 4.- Jorge Flores
- 5.- Cinthya Gomez
- 6.- Monserrat Gutierrez
- 7.- Jorgelina Tissera
- 8.- Cristian Falco
- 9.- Carlos Gimenez

Plan del Proyecto

Objetivo del Proyecto:

Evaluar y mejorar los modelos de aprendizaje automático para la predicción del churn (abandono de clientes) y aplicar estrategias basadas en clustering para mejorar la retención de clientes y optimizar las campañas de marketing.

Duración del Proyecto:

- Fecha de inicio: 22 de abril
- Fecha de finalización: 24 de mayo

Equipo del Proyecto:

- Walter Nieto Project Manager (PM) / ML Engineer
- Sol Figueroa ML Engineer /Analista de BI
- Jorgelina Tissera Analista de Datos
- Eduardo Figueroa Analistas de Datos
- Jorge Flores Analistas de Datos/ Analista BI
- Cristian Falco: Analista de Datos
- Carlos Giménez Analistas de Datos
- Cynthia Gomez Analistas de Datos/Calidad
- Monserrat Gutiérrez Analistas de Datos/ Estadística

Entregables del Proyecto:

1. Inicio del Proyecto (Semana 1)

- Kick-off Meeting: Reunión inicial para definir objetivos, roles y responsabilidades.
- Plan de Proyecto Detallado: Documento que incluye el cronograma del proyecto, los recursos asignados y el plan de comunicación.
- Revisión del Conjunto de Datos: Validación y limpieza inicial del conjunto de datos utilizado para la predicción del churn.

2. Análisis Exploratorio y Preprocesamiento (Semana 1 y 2)

- Análisis Exploratorio de Datos (EDA): Informe que detalla las características del conjunto de datos, identificando posibles problemas como datos faltantes y outliers.

- Preprocesamiento de Datos: Implementación de técnicas de normalización, imputación y codificación de datos, y generación de un informe de preprocesamiento.

3. Desarrollo de Modelos de ML (Semana 2 y 3)

- Entrenamiento de Modelos: Desarrollar y entrenar modelos de ML como Extra Trees Classifier, LightGBM, XGBoost y Random Forest.
- Evaluación de Modelos: Informe con métricas de rendimiento (precisión, AUC, recall, F1-score, kappa y MCC) para cada modelo.
- Selección del Mejor Modelo: Justificación de la selección del modelo óptimo basado en las métricas obtenidas.

4. Implementación de Técnicas de Clustering (Semana 3)

- Aplicación de Clustering: Implementar técnicas de clustering para segmentar a los clientes.
- Informe de Segmentación: Descripción de los clústeres identificados y sus características.

5. Validación y Optimización (Semana 3 y 4)

- Validación Cruzada: Realizar validación cruzada para asegurar la robustez de los modelos.
- Optimización de Hiperparámetros: Ajuste fino de los hiperparámetros de los modelos seleccionados.

6. Despliegue y Recomendaciones (Semana 4)

- Implementación del Modelo en Producción: Despliegue del modelo de ML y del sistema de clustering en un entorno de producción.
- Estrategias de Retención y Marketing: Informe con recomendaciones de estrategias basadas en los resultados de la segmentación para mejorar la retención y optimizar campañas de marketing.

7. Cierre del Proyecto (Final de Semana 4)

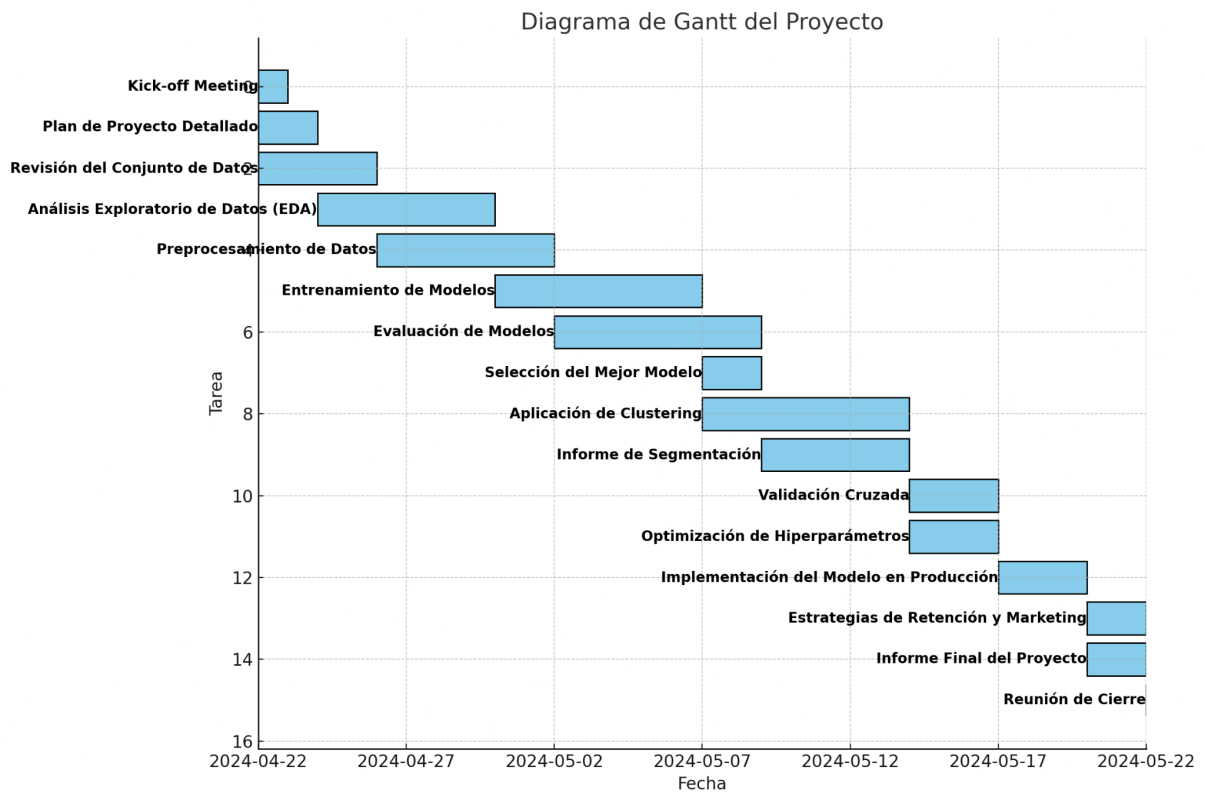
- Informe Final del Proyecto: Documento que resume los hallazgos, el rendimiento de los modelos, las recomendaciones y los pasos siguientes.
- Reunión de Cierre: Presentación de los resultados del proyecto a las partes interesadas y discusión de los siguientes pasos.

Cronograma del Proyecto:

Semana	Actividades Principales	Entregables
1	Inicio del proyecto, EDA, Planificación	Kick-off Meeting, Plan de Proyecto, EDA Report
2	Preprocesamiento de datos, Entrenamiento inicial	Informe de Preprocesamiento, Modelos Entrenados
3	Evaluación de modelos, Clustering	Informe de Evaluación, Informe de Clustering
4	Validación, Optimización, Despliegue	Informe de Validación y Optimización, Despliegue en Producción
Final	Cierre del proyecto	Informe Final, Reunión de Cierre

Notas Adicionales:

- **Herramientas y Tecnologías:** Se recomienda utilizar herramientas como Python, Jupyter Notebooks, librerías de ML (scikit-learn, XGBoost, LightGBM), y herramientas de visualización (Matplotlib, Seaborn y Power BI).
- **Comunicación:** Reuniones semanales de seguimiento y actualización de estado con el equipo y las partes interesadas.



El avance del proyecto está en Microsoft Project.