Guía Proyecto Final Grupal: Validación de conocimientos

***Machine Learning Operations Bootcamp - 2a edición***



[**INTRODUCCIÓN 2**](#_l3akqtxuz58e)

[**VALIDACIÓN DE CONOCIMIENTOS 3**](#_p4vn84n7a6uc)

[**EQUIPOS**](#_chgc7fwg1qx3) **8**

[**LISTADO DE DATA SETS 10**](#_nueze4hiiejo)

[**CHECKPOINTS 10**](#_i88nhut63ph5)

[**ITINERARIO 10**](#_zidpnjt4d2in)

[**RÚBRICA DE EVALUACIÓN 10**](#_xy3ivieoe6ro)

## INTRODUCCIÓN

Esta guía está diseñada para brindarles los detalles necesarios para llevar a cabo el proyecto final grupal de validación de conocimientos. Les solicitamos que revisen detenidamente la información proporcionada a continuación para asegurar el éxito de su proyecto.

Como parte del acompañamiento que les ofrecemos, hemos iniciado conversaciones con los miembros de cada equipo para facilitar el seguimiento de sus proyectos. Les sugerimos llevar a cabo la planeación dentro de las mismas y no en canales externos.

Además, les recordamos que tienen a su disposición los siguientes espacios para resolver preguntas y dudas específicas:

* Canal de Slack: #foro-mlops-bootcamp-2a-edición
* Mentorías 1:1: Consultar las indicaciones detalladas en la guía del estudiante
* Checkpoints en office hour: Detalles especificados en esta guía

Por último, les pedimos que administren cuidadosamente los tiempos y que consideren los requisitos planteados. ¡Les deseamos mucho éxito en la realización de su proyecto!

**Atentamente**

**Wizeline Academy**

## 

## VALIDACIÓN DE CONOCIMIENTOS

**Descripción del Problema:** La implementación de una estrategia efectiva de MLOps es crucial en entornos empresariales modernos en donde los modelos de Machine Learning juegan un papel significativo en la toma de decisiones y la optimización de procesos.

***Colaboren en equipo para desarrollar un modelo de negocio de MLOps, evalúen la eficacia y pertinencia de implementar esta estrategia para resolver un problema particular, con el propósito de alcanzar los siguientes objetivos:***  
  
**Objetivos:** Justificar, durante la presentación, que la estrategia presentada por el equipo cumple los siguiente:

1. **Evaluación de la Viabilidad:** Evaluen la aplicabilidad del modelo de aprendizaje automático en entornos empresariales reales, identificando y superando desafíos para garantizar una implementación exitosa.
2. **Confirmación de Relevancia y Accionabilidad:** Confirmen la relevancia del modelo de aprendizaje automático al alinearlo con las necesidades y objetivos específicos de la empresa, asegurando que proporcionen información práctica y accionable.
3. **Medición del Impacto en los KPI:** Cuantifiquen el impacto del modelo de aprendizaje automático en los indicadores clave de rendimiento (KPI), evaluando su contribución a la mejora de la eficiencia, precisión y otras métricas relevantes.
4. **Alineación con los Objetivos Estratégicos:** Verifiquen la alineación entre las iniciativas de aprendizaje automático y los objetivos empresariales más amplios, asegurándose de que apoyen las estrategias y prioridades de la organización.
5. **Gobernanza de Datos:** Establezcan políticas y procedimientos para garantizar la calidad, integridad y confiabilidad de los datos utilizados en los modelos de aprendizaje automático, así como para garantizar el cumplimiento de regulaciones y estándares de la industria.
6. **Colaboración y Comunicación entre Equipos:** Establezcan canales efectivos de comunicación y colaboración entre los diferentes equipos involucrados en el desarrollo, implementación y mantenimiento del modelo de aprendizaje automático, incluidos Data Engineers, Software Engineers, SRE y Data Scientists.
7. **Creación:**Creen un modelo base y nuevas versiones del mismo modelo, que incluyan cambios en las características o ajustes de hiper parámetros para la reproducibilidad y el seguimiento experimental, asegurando versionamiento de artefactos de datos. También pueden crear otro modelo después de tener el modelo base para poder hacer comparativos entre ambos. Agreguen las estrategias de testeo que se aplicarán a su pipeline de aprendizaje automático.
8. **Implementación:**Implementen un modelo en su entorno local utilizando contenedores y planifiquen el reentrenamiento, el drift, el redeploy, el escalado y el monitoreo. ***Orquesten*** y ***registren*** el pipeline de aprendizaje automático, y **realicen** el despliegue inicial.
9. **Escalabilidad y Flexibilidad:** Diseñen la arquitectura de la solución de MLOps de manera que sea escalable y flexible para adaptarse a cambios en los requisitos empresariales y tecnológicos a lo largo del tiempo.
10. **Documentación y Transferencia de Conocimiento:** Documenten adecuadamente todo el proceso de desarrollo, implementación y mantenimiento del modelo de aprendizaje automático, y faciliten la transferencia de conocimientos entre los miembros del equipo y las partes interesadas.

**Requisitos:**

Cada equipo debe:

1. **Tener una presentación que incluya:**

* **Descripción del negocio:** Descripción general del dataset con la cantidad de registros, la variable objetivo y variables.
* **Elección de modelo y justificación:** Explicación del por qué abordar este desafío es crucial para mejorar la eficiencia y la toma de decisiones en la empresa.
* **Herramientas utilizadas en el pipeline de MLOps:** Detalle de las herramientas específicas que se utilizarán, como sistemas de gestión de versiones de código y de datos, lenguajes de programación, infraestructura de computación, etc.
* **Diseño de arquitectura:** Diagramas de la solución de extremo a extremo que muestren la arquitectura de la solución, desde la adquisición de datos hasta la implementación y monitoreo de los modelos.
* Los beneficios de implementar esta solución son evidentes. Desde una perspectiva empresarial, ¿cuáles consideran que serán las ventajas de desarrollar esta solución utilizando un framework orientado a MLOps en lugar de solo un modelo aislado?

1. **Tener un repositorio** que contenga el código completo del proyecto, detallando las contribuciones individuales de cada miembro del equipo. Documentar de manera clara y concisa tanto la implementación de la solución como el proceso de resolución de problemas dentro de un marco de documentación estructurado.
2. **Aplicar** los conceptos aprendidos en el programa para el desarrollo integral de tareas cotidianas de aprendizaje automático.

**Consideraciones importantes:**

* **Roles dentro del equipo:** Cada integrante del equipo debe representar uno de los roles descritos a continuación y aportar a la presentación desde las responsabilidades asociadas a ese rol ***(las responsabilidades mostradas a continuación son sólo ejemplos y no son limitativas).***
* **Tiempo de intervención:** Cada integrante deberá participar equitativamente durante la presentación 10 minutos máximo por participante y 20 minutos de Q&A.

| **Rol** | **Responsabilidades** |
| --- | --- |
| **Data Engineer** | * Preprocesamiento de datos para garantizar su calidad y adecuación para los modelos de aprendizaje automático. * Desarrollo e implementación de pipelines de datos para la ingesta de los modelos. |
| **Software Engineer** | * Desarrollo de sistemas de gestión de versiones de código para garantizar la reproducibilidad y el control de cambios en los modelos. * Integración de los modelos en los sistemas empresariales existentes. |
| **Site Reliability Engineer** | * Diseño e implementación de infraestructura escalable y robusta para la ejecución y despliegue de modelos de aprendizaje automático. * Monitoreo y gestión de la infraestructura para garantizar su disponibilidad y rendimiento. |
| **Data Scientist** | * Desarrollo y entrenamiento de modelos de aprendizaje automático que aborden desafíos específicos del negocio. * Evaluación y selección de métricas relevantes para medir el rendimiento del modelo. |

## 

## 

## 

## 

## 

## 

## EQUIPOS

| Equipo 1  * **DE -** Ulises López Díaz * **DS -** Luis Enrique Moreno Mendieta * **SRE -** Alan Ernesto Ortega Muño * **SWE -** Christian Michel Curiel Anaya | **Equipo 2**   * **DE -** Ángel Isaí Zavala Cano * **DS -** Francisco Bueno * **SRE -** José David Santander Martínez * **SWE -** Aldo Humberto Arellano Romero |
| --- | --- |
| **Equipo 3**   * **DE -** José David Barrita López * **DS -** Jorge Ángel Ruiz Basurto * **SRE -** Iván Arturo Segura Ramírez * **SWE -** Alejandro Valderrama Silva | **Equipo 4**   * **DE -** Marina Estefanía García Casillas * **DS -** Carlos Andrés Tapasco Viera * **SRE -** Rogelio Martín Pérez Chávez * **SWE -** Daniela Martín Martínez |
| **Equipo 5**   * **DE -** Miguel Ángel Medina Ávila * **DS -** Edgar Correa Alpizar * **SRE -** Luis Ángel Hermenegildo Domínguez * **SWE -** Diego Edgar Juárez Elizalde | **Equipo 6**   * **DE -** Adolfo Ulises Castro Santamaría * **DS -** Elisa Guerrero * **SRE -** Luis Ernesto Román Chitala * **SWE -** Humberto Mendoza |

| **Equipo 7**   * **DE -** Adrián Trujillo Duron * **DS -** Ricardo Diaz Ochoa * **SRE -** Laura Valenzuela * **SWE -** Hugo Jovan Ramírez Cero | **Equipo 8**   * **DE -** Mauricio Fernando Bautista López * **DS -** Jesús Omar Rodríguez Bravo * **SRE -** Airain Alejandra Montoya Rodríguez * **SWE -** Rodolfo Martínez Vega |
| --- | --- |
| **Equipo 9**   * **DE -** Uriel Francisco Esparza Eguia * **DS -** Carlos Alexis Bañuelos Lepe * **SRE -** José Luis Aguayo Aguilar * **SWE -** Jaime Isaac Cortés Zapata | **Equipo 10**   * **DE -** Elmer Payro Costilla * **DS -** Xóchitl Alejandra Cedillo Padilla * **SRE -** Sócrates Gerardo Anchondo Rivera * **SWE -** Leonardo Andrés Valdez Dimas |
| **Equipo 11**   * **DE -** Héctor Segura Quintanilla * **DS -** Cristopher Trapero Orona * **SRE -** Daniel Humberto Vargas López * **SWE -** Saúl Enrique Mendoza Ruiz | **Equipo 12**   * **DE -** Carlos Ávila * **SRE -** Guillermo Manuel Galván Soltero * **SWE -** Carlos Enrique Lezama Jacinto * **DS -** NA |

## 

## LISTADO DE DATA SETS

**Equipos 1, 3, 5, 7, 9, 11:** [Compras en una tienda minorista en línea](https://www.kaggle.com/datasets/vijayuv/onlineretail) **Equipos 2, 4, 6, 8, 10, 12:** [Demanda de reservaciones en hoteles](https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand)

## 

## CHECKPOINTS

* Checkpoint 1: Jul 17, 2024
* Checkpoint 2: Aug 14, 2024
* Checkpoint 3: Sep 11, 2024

## ITINERARIO

* **Se compartirá más adelante**

## [RÚBRICA DE EVALUACIÓN](https://docs.google.com/document/d/12T3AAiB9YUrXwK_ucH974_WWlhVF-_OwUjPp-GVP0Xg/edit?usp=sharing)