

## **PROYECTO FINAL**

### **TERCERA ENTREGA**

El propósito de este proyecto es aplicar alguna metodología de ciencia de datos en conjunto con las diferentes técnicas y herramientas vistas durante el semestre para resolver una necesidad o aprovechar alguna oportunidad identificada en una organización de su elección. Se recomienda ampliamente que se intente realizar un acercamiento a esta organización dueña de la problemática y de los datos, particularmente con aquellos *stakeholders* quienes puedan direccionar lo mejor posible los requerimientos de la solución de ciencia de datos a desarrollar.

#### **OBJETIVOS DE LA TERCERA ENTREGA**

- Finalizar las etapas de modelado y evaluación, teniendo en cuenta la retroalimentación brindada durante la sustentación de la segunda entrega.
- Implementar la estrategia de despliegue del modelo / solución analítica, dependiendo de las necesidades establecidas para la organización.
- Preparar la presentación ejecutiva con los resultados más relevantes del proyecto.

#### **TERCER SPRINT DEL PROYECTO**

El tercer sprint del proyecto debe enfocarse en la consolidación de resultados de las etapas de modelado y evaluación, el despliegue de la solución analítica y la presentación de resultados obtenidos a los *stakeholders* de la organización y/o posibles inversionistas. **Tenga en cuenta que la meta final del proyecto no es la construcción de un modelo de machine learning per se, sino la construcción de una solución analítica que apoyará algún proceso o unidad de negocio dentro la organización.** Asegúrese de definir, si aún no lo ha hecho, cuál es ese proceso o unidad de negocio, cuál es la necesidad u oportunidad, cómo se integra la solución analítica, cómo será usada por los usuarios finales y cómo se medirá su impacto.

Se recomiendan al menos las siguientes reuniones de grupo:

- *Reunión de lanzamiento y planeación del sprint:* Para definir roles y forma de trabajo del grupo. Se genera una lluvia de ideas sobre las próximas actividades a desarrollar así como los criterios de aceptación de las mismas.
- *Reuniones de seguimiento:* Se recomienda mínimo una reunión de seguimiento semanal corta. También pueden ser correos de avance según lo defina el grupo. Pueden tener un tablero de control para mayor visibilidad del estado actual de las tareas. Herramientas como Trello, Jira o Project son las más populares.
- *Reunión de finalización:* Para consolidar el entregable y analizar los aspectos a mejorar para el tercer y último sprint.

No olvide consolidar todo el trabajo de modelado, evaluación y despliegue de la solución en el repositorio de GitHub. Se recomienda seguir buenas prácticas de versionamiento de documentación.

## ACTIVIDADES DEL SPRINT

En este sprint se realizará una tercera y última iteración de la metodología ASUM-DM, con énfasis en las etapas de modelado, evaluación, despliegue y retroalimentación. Dentro del entregable se debe incluir lo siguiente:

1. **[20%] Construcción y evaluación del modelo final:** Detalle el proceso de experimentación que siguió para construir y seleccionar el mejor modelo, así como las características técnicas de dicho modelo. Realice una evaluación **cuantitativa y cualitativa** acorde con el enfoque analítico seleccionado. Dando continuidad a las conclusiones del segundo entregable, **se espera que pueda profundizar en las preguntas:** ¿Qué condiciones considera que deberían tener los datos para obtener mejores resultados? Más datos, mejores características, etc. ¿El modelo obtenido es suficiente para soportar la necesidad u oportunidad de negocio identificada?
2. **[30%] Despliegue del modelo:** Implemente la estrategia de despliegue del modelo. Esta fase puede incluir la construcción de una API REST para el consumo del modelo mediante peticiones HTTP o de un proceso ETL para, por ejemplo, la generación de predicciones en batch. Construya una aplicación web o móvil

sencilla o un dashboard/reporte como herramienta a ser usada por los usuarios finales. **No olvide incluir un diagrama con la arquitectura de la solución y su integración con la infraestructura actual de la organización.**

3. **[40%] Presentación final:** Debe tener una duración total de **máximo 10 minutos** y deben participar todos los integrantes del equipo. La presentación **debe incluir los resultados más relevantes de TODAS las fases de la metodología usada** y no solo las del último entregable. Es importante que esté **dirigida a un público ejecutivo** (directores, jefes de área, inversionistas, etc.) por lo que debe evitar el uso de conceptos técnicos complejos y enfocarse en la necesidad u oportunidad, solución planteada a alto nivel, resultados más relevantes e impacto para el negocio. También puede ser importante mencionar otros aspectos como los retos o problemas más importantes que se tuvieron, la forma en la que se solucionaron y el trabajo futuro.
4. **[10%] Autoevaluación y evaluación grupal:** Cada integrante debe completar la autoevaluación y evaluación de sus compañeros. La nota otorgada en este punto corresponderá al **promedio de calificaciones otorgadas** a título personal y de parte de sus compañeros.

## ENTREGA Y EVALUACIÓN

- El proyecto se debe realizar en grupos de 2 o 3 estudiantes.
- Debe entregarse un documento en formato PDF en donde se sinteticen los resultados y ajustes solicitados por parte de los docentes respecto a los dos primeros entregables, seguido del desarrollo de cada uno de los puntos descritos previamente. Debe tener máximo 12 páginas (incluida tabla de contenido y página de presentación), a una columna y con letra Arial tamaño 12.
- Debe incluirse un repositorio de GitHub con todo el código fuente desarrollado para la exploración y preparación de datos, construcción del modelo, evaluación y despliegue. El repositorio debe estar debidamente documentado. Puede utilizar el archivo Readme para describir el contenido de los diferentes scripts que se hayan creado así como el detalle de las diferentes técnicas y herramientas utilizadas. Si se va a utilizar alguna herramienta no vista en clase, debe discutirse previamente con el docente. **Si lo considera, puede separar el código fuente en varios repositorios**, por

ejemplo, uno para todo lo concerniente a la exploración, preparación de datos y modelado, otro para la construcción de la API REST y otro para la aplicación web o móvil. Asegúrese de referenciarlos adecuadamente.

- La fecha máxima de entrega por Bloque Neón es el **miércoles 30 de noviembre a las 12:00 p.m.** y el día siguiente se realizará la presentación final.
- Recuerde que la presentación final no debe tomar más de 10 minutos. **Los recursos necesarios para realizar dicha presentación deberán ser incluidos en la entrega.** Se seleccionará aleatoriamente el orden de las presentaciones.
- Se penalizará a los grupos que no cumplan con las pautas de formato establecidas para el documento, presentación final y/o repositorios de código.