

## **Guía de trabajo Módulo 2 – Aprendizaje Supervisado**

### **1. Caso de estudio**

La agencia de marketing Sterling Cooper Advertising tiene en su planta de empleados alrededor de 4.000 personas directamente contratadas. Sin embargo, el departamento de recursos humanos ha reportado cifras preocupantes a la dirección de la agencia, indicando que cada año, alrededor del **15%** de sus empleados abandonan la empresa y necesitan ser reemplazados, en la mayoría de los casos, con muy poco tiempo para el proceso de selección y contratación. La dirección cree que este nivel de bajas (empleados que se marchan, ya sea por decisión propia o porque son despedidos) es perjudicial para la empresa, por las siguientes razones:

- Los proyectos de los antiguos empleados se retrasan, lo que dificulta el cumplimiento de los plazos, con la consiguiente pérdida de reputación entre sus clientes y socios.
- El departamento de recursos humanos requiere mucha inversión por los niveles de rotación, así que la mayoría de su personal está dedicado a tareas de reclutamiento de nuevo talento, haciendo más lento el proceso de desarrollo de otras áreas dentro del departamento dedicadas por ejemplo a la formación o bienestar de sus empleados.
- En la mayoría de los casos, hay que formar a los nuevos empleados para el puesto y/o darles tiempo para que se adapten a la cultura de la agencia.

Por esta razón, la dirección ha contratado a su equipo de consultores con el objetivo de **predecir oportunamente qué empleados podrían abandonar la organización**, de manera que sea posible implementar acciones preventivas orientadas a retener a la mayoría de quienes estén en riesgo.

Asimismo, desean identificar **los factores que más influyen en la decisión de renunciar**, con el fin de diseñar estrategias sostenibles a largo plazo que fortalezcan la permanencia del talento dentro de la empresa.

### **2. Datos disponibles**

Para abordar el problema descrito cuenta con las siguientes bases de datos:

- **general\_data:** Información general del empleado
- **employee\_survey\_data:** resultados de encuesta realizada a los empleados respecto a su nivel de satisfacción con su empleo actual
- **manager\_survey\_data:** resultados obtenidos por los empleados en su última evaluación de desempeño.
- **time\_work:** contiene el tiempo promedio de dedicación del empleado al día.

Se les proporciona un diccionario de variables para que pueda tener un mejor entendimiento de los datos disponibles para el desarrollo del caso de estudio (*datasets.zip*).

### **3. Desarrollo del caso de estudio**

El trabajo estará compuesto por dos entregas:

1. La primera entrega corresponde al desarrollo del caso de estudio aplicando alguno de los algoritmos introductorios (*Regresión Lineal Múltiple o Regresión Logística*). En esta entrega se propone abordar los siguientes aspectos:
  - a. Limpieza y transformación de los datos
  - b. Análisis exploratorio de los datos.
  - c. Preparación de los datos y selección de variables.
  - d. Aplicación de algoritmo.
  - e. Evaluación y análisis del modelo base.

***Fecha de entrega: 9 septiembre***

2. La segunda entrega corresponde a la aplicación de otros algoritmos de Machine Learning que permitan abordar el problema descrito. En esta entrega se propone abordar los siguientes aspectos:
  - a. Selección y aplicación de algoritmos
  - b. Comparación con modelo desarrollado en la entrega 1
  - c. Afinamiento de hiper-parámetros
  - d. Evaluación y análisis del mejor modelo
  - e. Conclusiones finales

***Fecha de entrega: 30 septiembre (sustentación)***

Consideraciones importantes:

- La entrega debe realizarse en notebooks, el equipo es libre de ordenar el desarrollo del caso de estudio como mejor lo considere. Los notebooks generados deben estar organizados y documentados con los análisis realizados y las principales conclusiones. Si en los notebooks no se detalla el proceso de análisis y desarrollo, la nota de la entrega se verá afectada.
- Se debe generar un informe donde se detalle la solución desarrollada, los análisis realizados, las conclusiones y recomendaciones del trabajo. El informe debe tener máximo 5 páginas.

### **4. Criterios de evaluación:**

- Pertinencia y creatividad de la solución entregada 5%
- Análisis exploratorio realizado a los datos disponible 20%
- Estructura, documentación y justificación del código 30%
- Análisis de resultados 35%.
- Redacción adecuada del informe (incluye ortografía) 10%

**5. Sustentación:**

- Cada grupo debe sustentar su entrega (*día de sustentaciones: 30 de septiembre*). Si el equipo no puede sustentarlo la nota será cero independientemente de la solución entregada.
- La presentación es de **máximo 15 minutos**.
- Se puede presentar en notebook o en presentaciones elaboradas por el equipo.

**Los criterios de evaluación son:**

- Claridad y orden en la presentación (presentar lento y escoger bien los contenidos a mostrar).
- Compresión adecuada del problema.
- No tener errores conceptuales (mal uso o mala interpretación de conceptos o herramientas).