

ESCUELA DE TALENTO

DIGITAL

- 100% ONLINE ■ MENTORIZACIÓN PERMANENTE
- ORIENTADO A LA EMPLEABILIDAD ■ GRATUITO
- CONEXIÓN CON EL MERCADO

NTT DATA FOUNDATION

ESCUELA DE TALENTO DIGITAL

NTT DATA FOUNDATION

RETO GRUPAL ÁREA 3

ÍNDICE

1. INTRODUCCIÓN	3
1.1. ¿Cómo entregáis vuestros ejercicios?	3
1.2. ¿Qué debe contener el documento pdf?	3
2. ESTADOS DE ESTADOS UNIDOS	3
2.1. Ejercicio 1	4
2.2. Ejercicio 2	5
2.3. Ejercicio 3	5
2.4. Ejercicio 4	6

1. INTRODUCCIÓN

Para superar este reto grupal, tendréis que ir resolviendo una serie de ejercicios que os vamos a proponer en este documento.

1.1. ¿Cómo entregáis vuestros ejercicios?

Tendréis que preparar un documento pdf y subirlo a la plataforma en el espacio habilitado para ello. No es necesario que todos los componentes del grupo subáis el documento, con que lo suba uno de vosotros es suficiente.

1.2. ¿Qué debe contener el documento pdf?

Este documento deberá contener la propuesta de solución que resuelva cada uno de los ejercicios, pero, además, también debe contener una explicación de cómo habéis llegado a obtener esa solución, que debe ser conjunta y aprobada por todos los miembros del grupo.

Como durante el desarrollo de la actividad van a surgir diferentes propuestas, queremos que las documentéis, es decir, que cuando expliquéis cómo habéis llegado al resultado final, también tenéis que explicar qué otras alternativas había y quién las ha propuesto.

Por ejemplo, imaginad un grupo de 5 alumnos (alumno1, alumno2, alumno3, alumno4 y alumno5) resolviendo el ejercicio 1. La propuesta de resolución del ejercicio 1 debería ser algo como esto:

Después de leer el enunciado, entendimos que lo que se solicitaba era hacer

Durante el proceso, el estudiante1 propuso llegar a la solución de la siguiente manera..... pero al estudiante2 y al estudiante3 les pareció mejor hacer y todos estuvimos de acuerdo.

Por todo esto, proponemos esta solución en la que estamos de acuerdo los 5 participantes:

SOLUCIÓN PROPUESTA

Con esto lo que queremos valorar es la participación de cada uno de vosotros durante el desarrollo del reto.

Cuidad también el formato en el que presentéis el documento, porque también se tendrá en cuenta.

Si tenéis cualquier duda, consultad al tutor a través de la plataforma.

2. ESTADOS DE ESTADOS UNIDOS

Después de haber trabajado en los dos primeros retos, ya sabéis algo más de los habitantes de los Estados de Estados Unidos. En este reto, vamos a profundizar un poco

más en cómo se podrían comportar las personas de tres de estos estados, en relación con su banco.

En concreto, una entidad bancaria americana que opera en California, Texas y Alabama pretende implementar un modelo de abandono de clientes (modelo Churn) con el objetivo de poder anticiparse a la marcha de estos y poder realizar acciones comerciales para, precisamente, evitar la pérdida implícita en el abandono de los clientes.



La entidad bancaria, nos facilita una base de datos con 10.000 registros de información que contienen las siguientes variables:

- **RowNumber**: número de registro.
- **CustomerId**: identificador único de cliente.
- **Surname**: apellido del cliente.
- **CreditScore**: puntuación de crédito del cliente.
- **Geography**: estado del cliente.
- **Gender**: género del cliente.
- **Age**: edad del cliente.
- **Tenure**: número de años que el cliente lleva en el banco.
- **Balance**: balance del cliente.
- **NumOfProducts**: número de productos que el cliente tiene en el banco.
- **HasCrCard**: si el cliente tiene o no tarjeta de crédito.
- **IsActiveMember**: es cliente activo o no de la entidad.
- **EstimatedSalary**: salario que la entidad estima que ingresa mensualmente el cliente.
- **Exited**: es la variable objetivo:
 - **1**: el cliente abandonó la entidad
 - **0**: el cliente no abandonó la entidad

2.1. Ejercicio 1

Leed el fichero de texto [df_estados_bank.csv](#) que se corresponde con el conjunto de datos a trabajar en este reto.

Realizad una correcta **limpieza de datos inicial** con el objetivo de preparar la base de datos previa a entrenar el modelo. **Sin haber entrenado el modelo**, dad respuesta a las siguientes preguntas:

- ¿Consideráis que todas las variables del conjunto de datos son útiles de cara a la predicción? ¿Cuáles sí y cuáles no? Justificad vuestra respuesta.
- ¿A qué tipo de modelo os enfrentáis? Justificad vuestra respuesta.

2.2. Ejercicio 2

Atendiendo a la segunda respuesta del ejercicio anterior, **plantead**, al menos, **2 algoritmos diferentes que predigan el abandono** (o no) del cliente. Realizad todos los pasos necesarios antes de la aplicación del modelo predictivo.



- ¿Qué algoritmo es el que os ha arrojado mejores resultados?
- ¿Cómo habéis evaluado la bondad del ajuste? Justificad vuestra respuesta.
- ¿Podéis explicar al cliente qué características son las que más inciden en la decisión de un cliente para abandonar definitivamente la entidad?

2.3. Ejercicio 3

Escogiendo el mejor algoritmo del ejercicio anterior, planteaos la posibilidad de que los resultados puedan ser más ajustados si utilizáis **un modelo diferente por cada estado** en que la entidad opera.



- ¿Tenéis algún problema de **desbalanceo** de datos? Si es así, solucionadlo.
- ¿Obtenéis mejores resultados planteando modelos diferenciados por estado?

2.4. Ejercicio 4

Plantead un algoritmo de clusterización en el que se pretenda encontrar grupos de clientes homogéneos.

- ¿Qué características tienen los clientes de cada uno de los grupos?