

Tarea:

Análisis descriptivo del conjunto de datos Bank Marketing: Caracterización de clientes y evaluación de campañas

1. Objetivo del trabajo

El objetivo de este informe es explorar y analizar de forma descriptiva el conjunto de datos Bank Marketing, el cual contiene información sobre campañas de marketing directo de un banco portugués. A través del análisis estadístico de variables demográficas, financieras y de contacto, se busca comprender las características principales de los clientes y los factores asociados al éxito de las campañas, identificando patrones relevantes en la variable objetivo: la suscripción a depósitos a plazo.

2. Descripción del Conjunto de Datos Bank Marketing

Información Básica

- **Nombre del conjunto de datos:** Bank Marketing
- **Formato:** CSV (Comma-Separated Values) y XLSX
- **Cantidad de registros:** 41188 clientes
- **Cantidad de variables:** 21 variables

Contexto General

Este conjunto de datos contiene información sobre campañas de marketing directo de un banco portugués. Los datos fueron recolectados principalmente mediante llamadas telefónicas con el objetivo de determinar si los clientes suscribirían un depósito a plazo bancario.

La campaña de marketing se realizó entre mayo de 2008 y noviembre de 2010, abarcando el periodo de la crisis financiera global. El banco contactó a sus clientes para ofrecerles productos financieros, específicamente depósitos a plazo, que podrían proporcionar intereses más seguros durante un periodo económicamente inestable.

Variables Incluidas

Las 21 variables del conjunto se pueden categorizar en:

- 1. Datos demográficos del cliente (5 variables):**

edad, trabajo, estado civil, educación, si tiene préstamo impagado

- 2. Situación financiera del cliente (2 variables):**

si tiene préstamo hipotecario, si tiene préstamo personal

- 3. Información de contacto (4 variables):**

tipo de contacto, mes, día de la semana, duración de la llamada

- 4. Historial de campañas (4 variables):**

número de contactos durante esta campaña, días desde último contacto, contactos previos, resultado de campañas anteriores

- 5. Indicadores económicos (5 variables):**

tasa de variación del empleo, índice de precios al consumidor, índice de confianza del consumidor, tasa Euribor a 3 meses, número de empleados

- 6. Variable objetivo (1 variable):**

si el cliente suscribió un depósito a plazo ("yes" o "no")

Propósito de la Recolección

Los datos fueron recolectados por la institución bancaria con varios objetivos:

1. Mejorar la eficiencia de sus campañas de marketing directo
2. Identificar qué clientes tienen mayor probabilidad de suscribir depósitos a plazo
3. Comprender qué factores (personales, de contacto o económicos) influyen en la decisión del cliente
4. Optimizar el uso de recursos en futuras campañas de marketing

Este conjunto de datos se ha convertido en un recurso estándar para investigaciones en el campo del aprendizaje automático y análisis predictivo, especialmente para modelado de clasificación binaria, siendo utilizado frecuentemente en entornos académicos y profesionales para desarrollar y probar algoritmos de análisis de datos.

Variables del conjunto de datos Bank Marketing (al español)

A continuación, se presenta el nombre de cada variable tal como aparece en la base de datos:

1. **age:** Edad del cliente
2. **job:** Tipo de trabajo u ocupación del cliente
3. **marital:** Estado civil
4. **education:** Nivel educativo
5. **default:** Si tiene crédito en incumplimiento (moroso)
6. **housing:** Si tiene préstamo hipotecario
7. **loan:** Si tiene préstamo personal
8. **contact:** Tipo de comunicación para el contacto (celular o teléfono fijo)
9. **month:** Mes del año en que se realizó el último contacto
10. **day_of_week:** Día de la semana en que se realizó el último contacto
11. **duration:** Duración del último contacto en segundos
12. **campaign:** Número de contactos realizados durante esta campaña con este cliente
13. **pdays:** Días transcurridos desde el último contacto en una campaña anterior
14. **previous:** Número de contactos realizados con este cliente antes de esta campaña
15. **poutcome:** Resultado de la campaña de marketing anterior
16. **emp.var.rate:** Tasa de variación de empleo (indicador económico trimestral)
17. **cons.price.idx:** Índice de precios al consumidor (indicador económico mensual)
18. **cons.conf.idx:** Índice de confianza del consumidor (indicador económico mensual)
19. **euribor3m:** Tasa Euribor a 3 meses (indicador económico diario)
20. **nr.employed:** Número de empleados (indicador económico trimestral)
21. **y:** Variable objetivo - si el cliente suscribió un depósito a plazo (sí o no)

Instrucciones

Responda las siguientes preguntas a través del desarrollo de un informe, elaborado en grupos de no más de 5 integrantes. Recuerde que el enfoque para responder las preguntas planteadas es desde el punto de vista descriptivo (exploratorio) revisado en clases.

El informe (en formato PDF) y código PYTHON, debe ser entregado en FECHA Y HORA DEADLINE, a través de la plataforma webcursos.

Se entrega un informe por equipo de trabajo.

Fecha de entrega: 20 de Julio, 23:55 horas.

Preguntas del Análisis

1. ¿Cuáles son las características principales de la población y muestra del conjunto de datos Bank Marketing, qué tipos de variables contiene y cómo está distribuida la variable objetivo?
2. ¿Cuál es la distribución de los clientes según su tipo de trabajo (job)?
3. ¿Cómo se distribuye el resultado de la campaña (y) según los meses del año (month)?
4. ¿Cuál es la distribución de edades (age) de los clientes contactados?
5. ¿Cómo se distribuye la duración de las llamadas (duration) realizadas durante la campaña?
6. ¿Cómo varía la distribución de la edad de los clientes (age) según su estado civil (marital)?, ¿existen diferencias significativas en la tendencia central, dispersión y presencia de valores atípicos entre los distintos grupos?
7. ¿De qué manera se distribuye la duración de las llamadas (duration) en función del resultado de la campaña (y)?, ¿se pueden observar diferencias en los patrones de tiempo invertido entre las llamadas que resultaron exitosas (suscripción) y las que no?
8. ¿Qué relación existe entre la edad del cliente (age) y la duración de la llamada (duration), y cómo se distribuyen estos patrones cuando se clasifican según si el cliente finalmente suscribió un depósito a plazo (y)?
9. ¿Qué variables numéricas del conjunto de datos Bank Marketing presentan las correlaciones más significativas entre sí, y cómo se podría reducir la

dimensionalidad del modelo predictivo basándose en estas correlaciones para evitar redundancias sin perder poder explicativo?

Cada una de las respuestas de cada una de las preguntas debe contener resultados, interpretación, discusión (de ser necesario haga supuestos) y conclusión.

Criterios de uso responsable de Inteligencia Artificial

En la elaboración de este trabajo, se permite el uso de herramientas basadas en inteligencia artificial como apoyo para el análisis de datos, en la misma línea en que se utilizan softwares estadísticos u otras tecnologías analíticas. No obstante, es importante tener presente las siguientes consideraciones:

- La responsabilidad sobre el contenido, análisis e interpretaciones presentadas recae exclusivamente en los estudiantes.
- Cada resultado debe ser verificado y validado por quienes elaboran el informe, garantizando su precisión y pertinencia dentro del contexto del análisis propuesto.
- El desarrollo del código y del informe debe ser escrito y comprendido por cada integrante del equipo. Toda solución debe ser justificada, demostrando el entendimiento del proceso seguido y de los resultados obtenidos.
- El uso de herramientas tecnológicas no exime al estudiante de su responsabilidad académica y profesional respecto a la calidad, veracidad y originalidad del trabajo presentado.
- En caso de utilizar herramientas de inteligencia artificial, se deberá indicar explícitamente en qué parte del trabajo fueron empleadas e incluir el o los prompts utilizados. Esta información puede ser incorporada como un anexo o agregada al final del documento.

La omisión de esta declaración, en caso de evidenciarse uso no informado de IA, será considerada una falta a la integridad académica.