

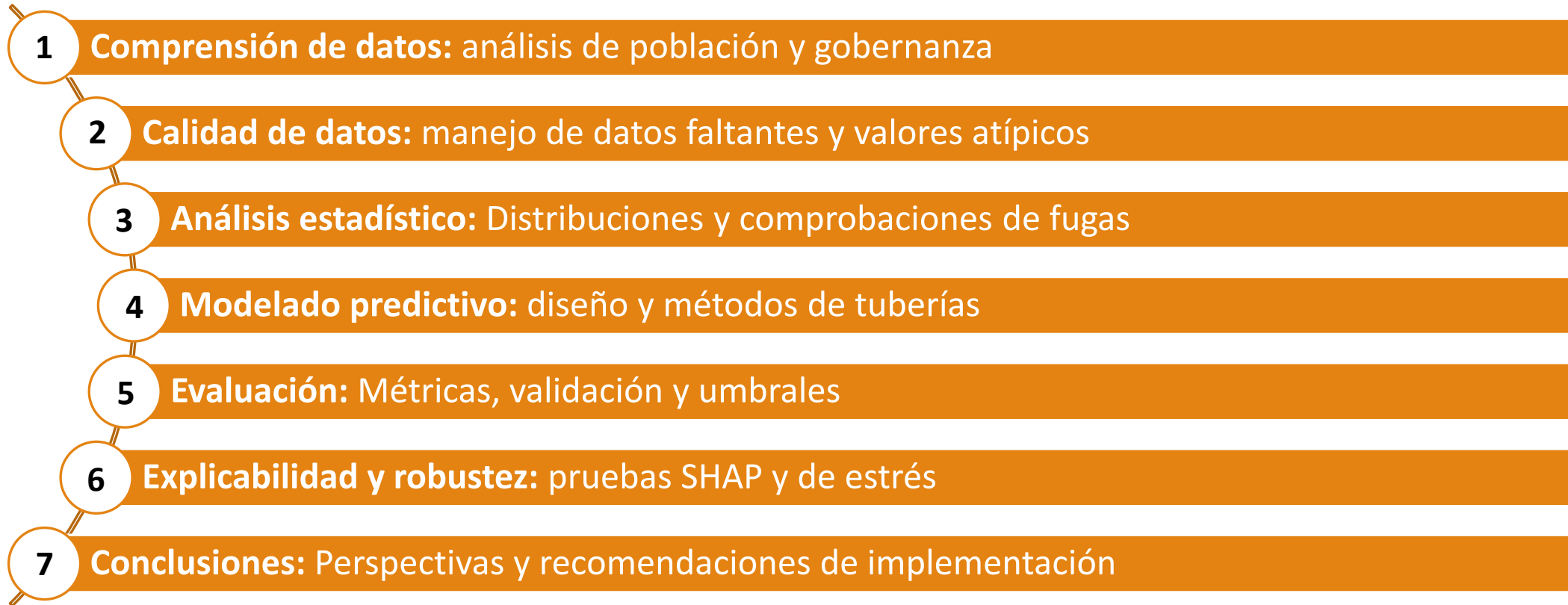
Modelado predictivo y análisis de datos de campañas de marketing bancarias



ANÁLISIS DE LOS DATOS DE CONTACTO DE CLIENTES DEL CONJUNTO DE DATOS DE MARKETING DEL BANCO PORTUGUÉS UCI PARA PREDECIR LOS RESULTADOS DE LAS SUSCRIPCIONES DE MANERA EFECTIVA Y CON RIGOR ESTADÍSTICO.

Hoja de ruta de análisis y modelado de datos

Enfoque gradual que abarca datos, modelado y evaluación



Descripción general del análisis descriptivo y la gobernanza de experimentos

Descripción general de las estrategias de población, modelado y partición de datos



Objetivo: Predecir la probabilidad de suscripción
 $P(y=\text{"sí"} \mid x)$

Pronosticar la probabilidad de suscripción del cliente



Unidad de análisis: Instancia de contacto del cliente

Cada contacto como unidad analítica



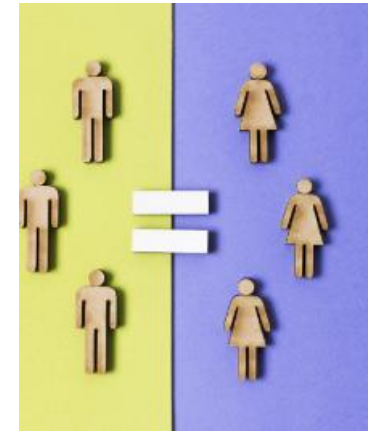
Caso de uso: modelos separados previos y posteriores al contacto

El precontacto excluye la "duración" para evitar fugas



Particionado de datos: división por tiempo y mes

Entrenar, validar y probar en meses consecutivos



Objetivo: Reducir el sesgo y medir la generalización temporal

Garantizar un rendimiento realista del modelo

Estrategias de codificación y calidad de datos en el marketing bancario

Manejo de datos faltantes, valores atípicos y codificación categórica

Aspecto	Detalles	Enfoque
Valores Faltantes / "Unknown"	Campos: resultados, educación, trabajo	Cuantificar y considerar el modelado como categoría explícita
Valores atípicos	Numéricos: edad, saldo, duración, campaña, pdays, anterior	Detectar mediante percentiles robustos (p1–p99); aplicar winsorización si es necesario
Cardinalidad categórica	Baja cardinalidad: trabajo, pareja, defecto...	Codificación one-hot con handle_unknown='ignore'
Categóricos de cardinalidad media	Algunas variables con categorías moderadas	Codificación de destino con pliegues de validación cruzada, suavizado interno para evitar fugas
Integridad de los datos	Garantizar la integridad y la coherencia	Validación periódica y detección de anomalías
Fundamento de la estrategia de codificación	Equilibrio entre la complejidad del modelo y el sesgo	Utilice técnicas de codificación adecuadas a los niveles de cardinalidad
Impacto atípico	Posible distorsión del rendimiento del modelo	Winsorizar para limitar los valores extremos

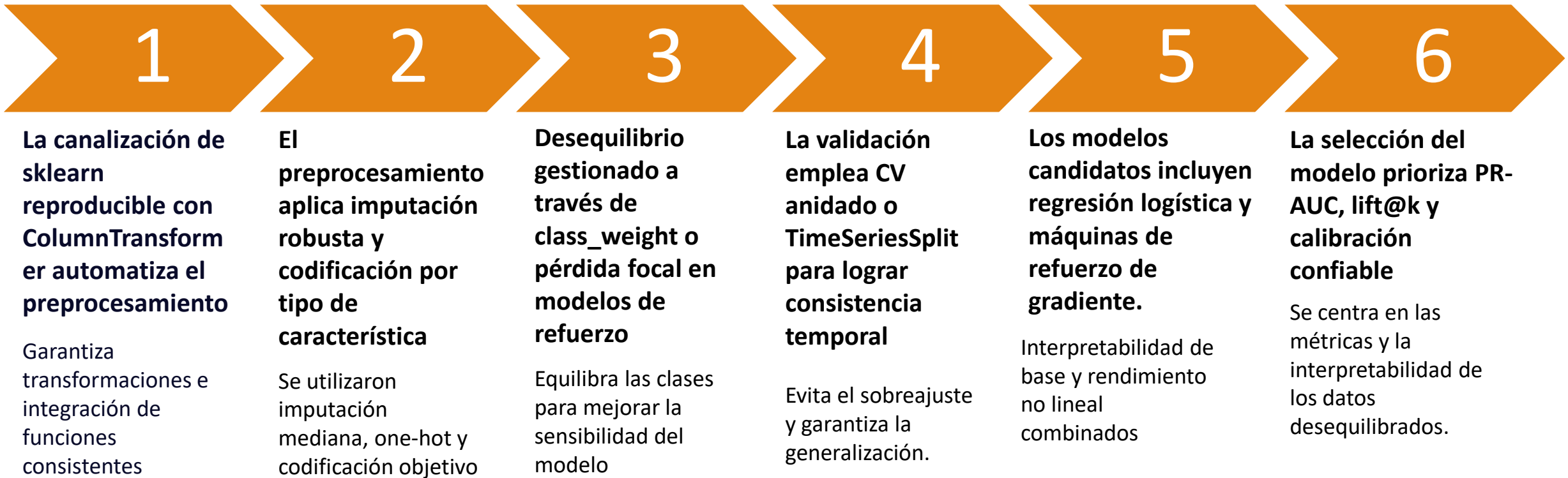
Descripción general del análisis univariado y bivariado

Distribuciones, asociaciones y controles de fugas en el marketing bancario

	Métricas	Objetivo
Univariante	Gráficos de densidad, histogramas, tablas de frecuencia	Examinar distribuciones y desequilibrio de clases (tasa base $\pi = P(y = \text{sí})$)
Bivariado (categórico vs. objetivo)	Odds Ratio, Peso de la Evidencia (PDE), Valor de la Información (VI), Prueba χ^2 , V de Cramér	Evaluar la fuerza de la asociación y el poder predictivo
Bivariado (Numérico vs. Objetivo)	Binning monótono (cuantiles), tasas de conversión por bin, correlación puntual-biserial, AUC univariante, estadística KS	Comprobar la monotonía y la capacidad de discriminación
Numérico vs numérico	Correlación de Spearman	Evaluar las relaciones monótonas
Control de fugas	Variable de duración excluida antes del contacto (fuga conocida)	Revisar las variables derivadas (poutcome, pdays) para detectar sesgo de actualidad y estabilidad

Justificación de la metodología y canalización de modelado predictivo

Enfoque estructurado para una predicción robusta del marketing bancario



Configuraciones del modelo: casos de uso previos y posteriores al contacto

Análisis comparativo de estrategias de modelado y uso de características

Aspect	Pre-contact Model	Post-contact Model
Caso de uso	Segmentación antes de la llamada	Modelado después del contacto
Inclusión de características clave	Excluye 'duración' (riesgo de fuga)	Incluye 'duración' (KPI posterior a la llamada)
Variable objetivo	Probabilidad de suscripción antes de la llamada	Probabilidad de suscripción post-contacto
Ingeniería de características	Excluir variables relacionadas con los resultados de llamadas con riesgo de fuga	Incluir variables de resultados de llamadas teniendo en cuenta el sesgo de actualidad
Estrategia de formación	División consciente del tiempo, enfoque en la generalización	Lo mismo, pero con funciones de datos de llamadas adicionales
Métricas de evaluación	Métricas de clasificación priorizadas (PR-AUC, Lift@k)	Métricas de calibración importantes (Brier, curvas de calibración)

Métricas de clasificación clave y métodos de validación

Descripción general de métricas y estrategias para la evaluación y robustez de modelos

Métricas	objetivo	Detalle
ROC-AUC	Discriminación	Área bajo la curva ROC; rendimiento de clasificación general
PR-AUC	Clasificación en datos desequilibrados	Más informativo cuando la clase positiva es rara
KS Statistic	Separación	Diferencia máxima entre distribuciones de puntuaciones de clases
Brier Score	Calidad de calibración	Mide la diferencia cuadrática media entre las probabilidades previstas y los resultados

Evaluación de la explicabilidad, estabilidad y robustez del modelo

Evaluaciones clave sobre explicabilidad, estabilidad, calibración y pruebas de robustez



Explicabilidad utilizando valores SHAP a niveles de observación global e individual

Los valores SHAP revelan las principales características predictivas e incluyen dependencia parcial y gráficos ICE para la visualización del impacto de las características.



Estabilidad evaluada comparando la importancia de las características y el índice de estabilidad de la población

La importancia de las características se compara entre pliegues y meses; PSI detecta la desviación de datos entre los conjuntos de entrenamiento y de prueba.



Calibración evaluada mediante gráficos de confiabilidad y métricas de error de calibración

Los gráficos de confiabilidad junto con los errores de calibración esperados y máximos cuantifican la precisión de la predicción probabilística del modelo.



Robustez probada con ruido, escenarios hipotéticos, ablación y controles de regularización

Las pruebas de estrés incluyen ruido de entrada y cambios de escenario; los estudios de ablación y la regularización de coeficientes verifican la resiliencia del modelo.

Adoptemos estrategias predictivas avanzadas para maximizar el impacto del marketing y garantizar modelos sólidos y transparentes.

Implemente modelado dual, preprocesamiento robusto, métricas de clasificación y herramientas de explicabilidad para mejorar los resultados de las campañas de marketing bancarias y la confiabilidad del modelo.