

Lead Scoring: Priorizando a tus futuros clientes

Descubre cómo clasificar y priorizar prospectos para optimizar tu proceso de ventas y marketing digital. Una metodología basada en datos para aumentar tu tasa de conversión.

D por Dani Castillo Cabrera





Beneficios del Lead Scoring



Priorización efectiva

El equipo de ventas se centra en los prospectos más valiosos, optimizando recursos y tiempo.



Eficacia en ventas

El proceso se vuelve más eficiente al contactar a los leads cuando están "calientes".



Personalización

Permite adaptar ofertas y mensajes según el perfil y comportamiento de cada prospecto.



Incremento en ventas

La estrategia focalizada resulta en mayor número de conversiones y mejor ROI.



Configuración del Proyecto

9,100

Registros iniciales

Base de datos completa para el análisis

20

Variables

Parámetros para evaluación de prospectos

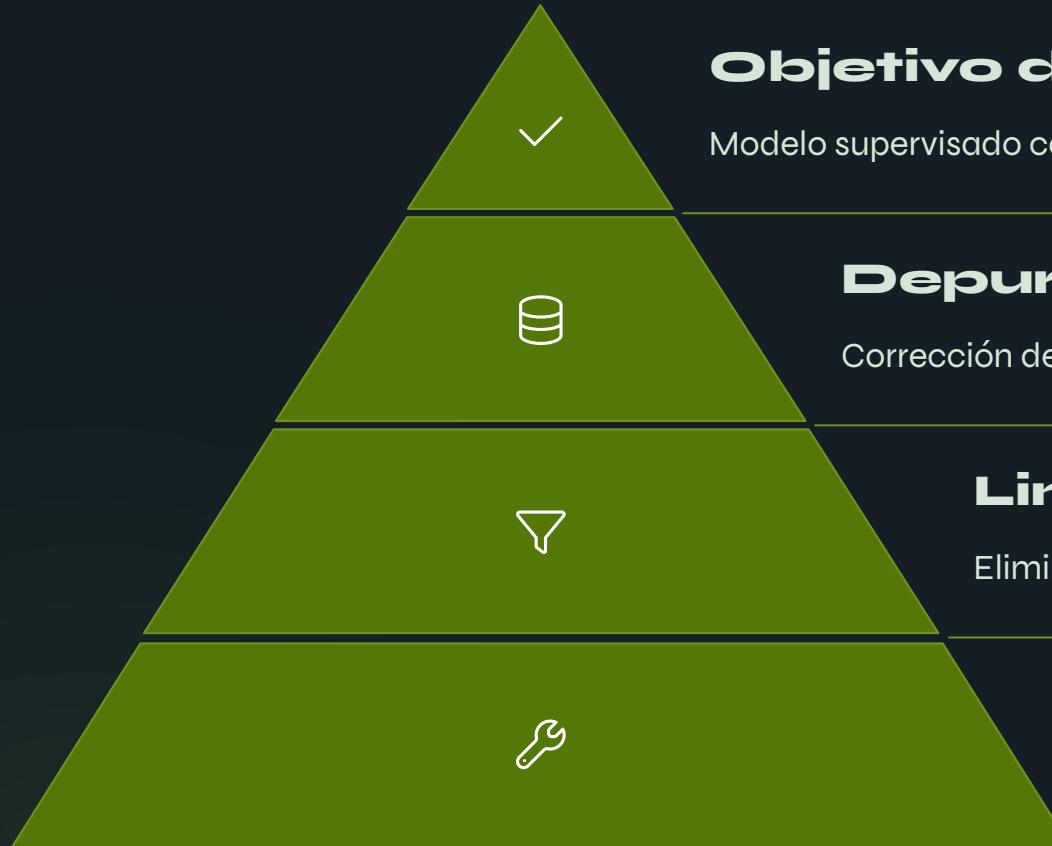
20%

Set de validación

Datos reservados para verificar el modelo

Iniciamos con un conjunto de datos robusto para garantizar la fiabilidad del modelo de scoring y su correcta validación.

Análisis de Datos



La preparación meticulosa del dataset garantiza que el modelo se construya sobre información fiable y relevante.

Análisis Exploratorio de Datos

Filtrado selectivo

Eliminación de variables irrelevantes y registros con emails falsos

Selección final

11 variables categóricas y 5 numéricas para el modelo



Origen de prospectos

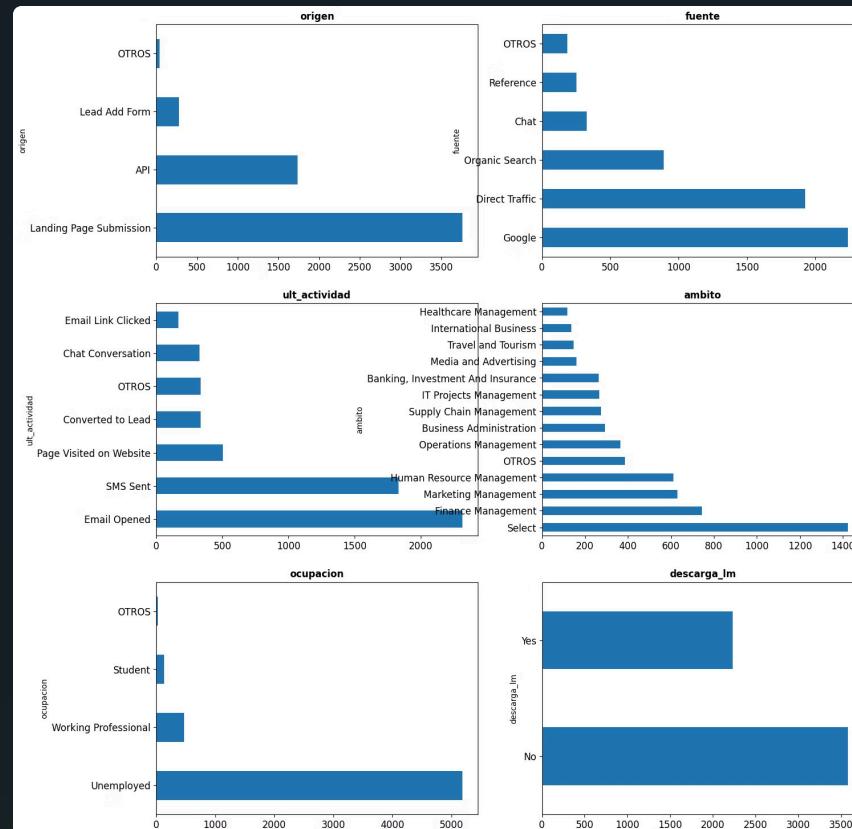
Identificación de landing pages como principal fuente

Comportamiento

Análisis de apertura de SMS y emails como indicadores clave

El análisis profundo nos permitió entender los patrones de comportamiento de los leads y seleccionar las variables más predictivas.

Visualización de Patrones

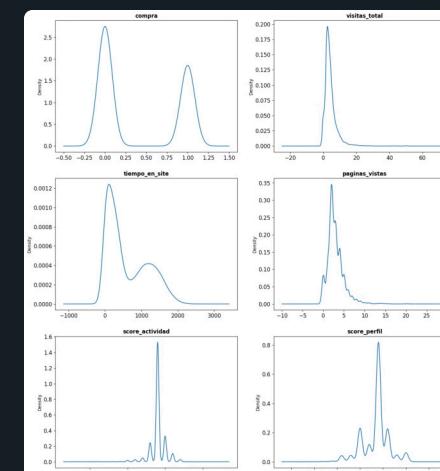


Distribución por fuente

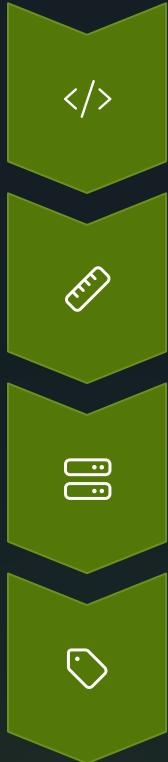
El gráfico muestra la prevalencia de leads provenientes de landing pages específicas. Estas fuentes generan prospectos con mayor potencial de conversión.

Comportamiento por canal

Los datos revelan patrones distintivos en la interacción con diferentes canales de comunicación. La respuesta a emails muestra correlación con la intención de compra.



Transformación de Datos



One-Hot Encoding

Transformación de variables categóricas a formato numérico

Min-Max Scaling

Normalización de variables numéricas entre 0 y 1

Unificación

Combinación de todas las variables transformadas

Separación objetivo

Aislamiento de la variable "COMPRA" para entrenamiento

La correcta transformación de datos prepara el terreno para que los algoritmos puedan aprender eficazmente de la información disponible.



Selección de Variables

Revisión visual

Evaluación preliminar de la relevancia de cada variable mediante análisis de distribución y correlación con la variable objetivo.

Recursive Feature Elimination

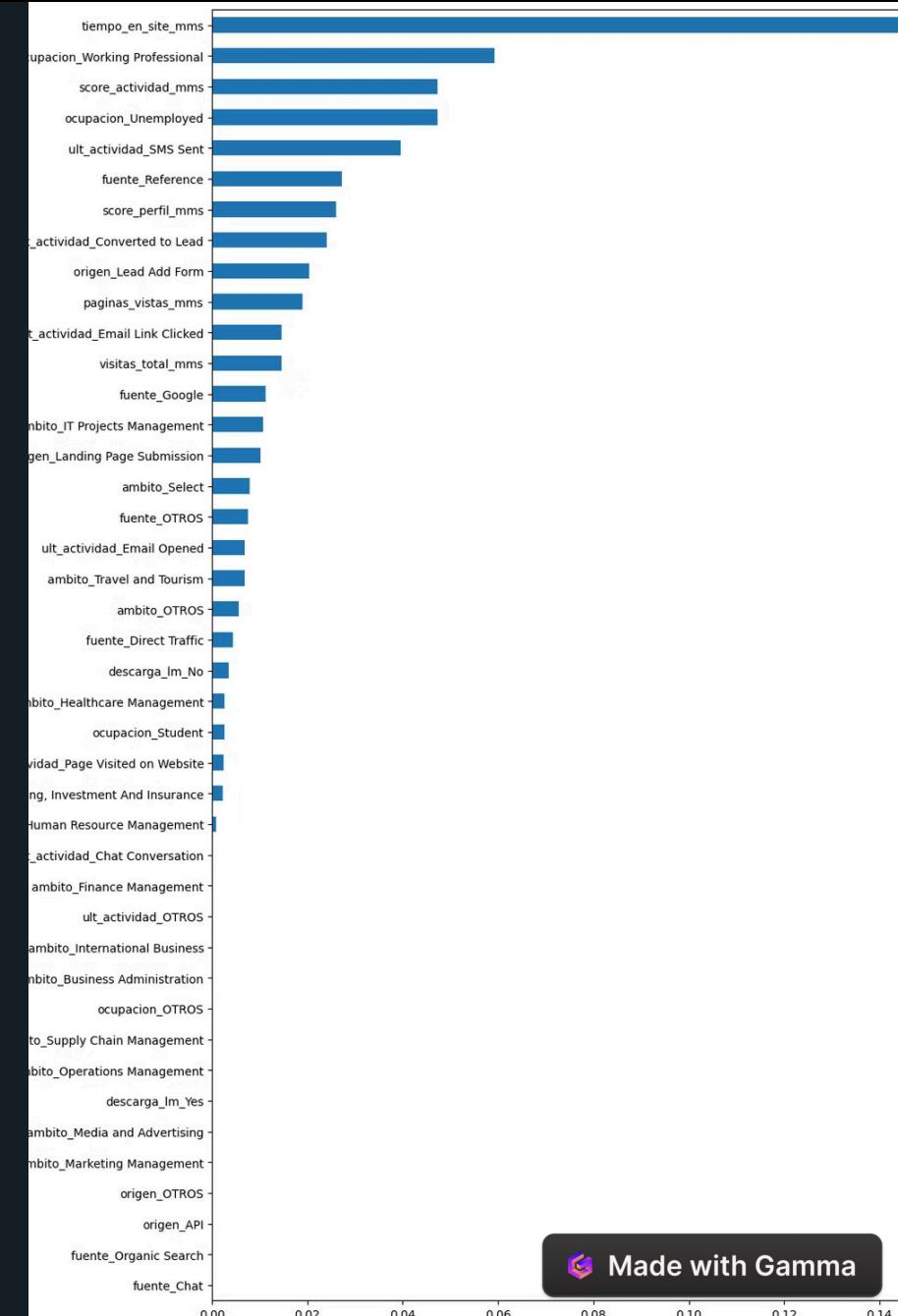
Aplicación de técnica iterativa para descartar variables de menor impacto en la capacidad predictiva del modelo.

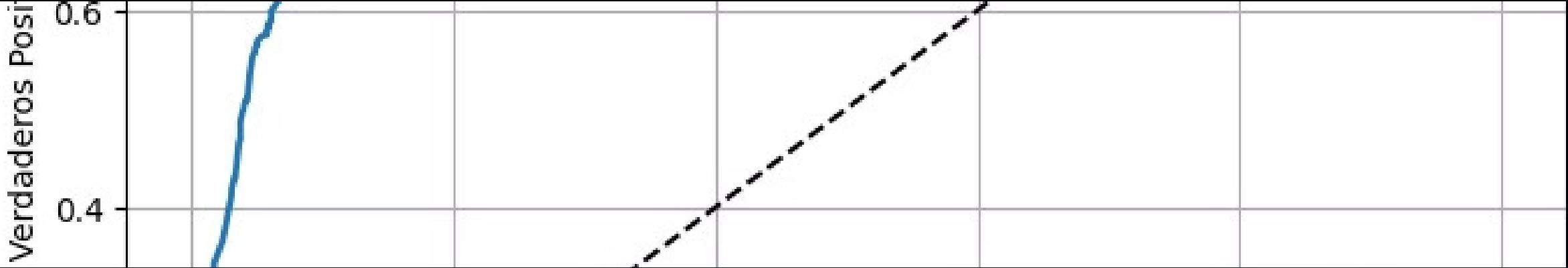
Permutation Importance

Método final seleccionado para determinar la importancia de cada variable mediante la medición del impacto en el rendimiento.

Análisis de correlación

Identificación y eliminación de variables con alta redundancia para evitar multicolinealidad en el modelo final.





Modelización

Modelo	AUC	Características
XGBoost	0.839	Alto rendimiento pero mayor complejidad
Random Forest	0.884	Excelente equilibrio entre precisión y estabilidad
Regresión Logística	0.874	Buena interpretabilidad pero menor capacidad predictiva

El modelo Random Forest demostró superioridad en rendimiento y estabilidad tras optimizar hiperparámetros mediante Grid Search.

Draws of Data Pipeline

Some pipelines are designed to connect to third-party systems like the Connect and the Pipelines, others connect to a clearing system or a clearing system. Mental calculations data from automated systems, enables you to evaluate and plan your plans.



Implementación en Producción

- 1 Selección final de variables**
Utilización exclusiva de los predictores más relevantes identificados.
 - 2 Estructuración del dataset**
Eliminación de duplicados y registros irrelevantes para optimizar el rendimiento.
 - 3 Creación de pipeline**
Desarrollo de flujo automatizado para tratamiento de valores nulos y atípicos.
 - 4 Implementación del sistema**
Integración del pipeline final de ejecución en la infraestructura existente.
- La implementación en producción garantiza que el modelo funcione de manera automatizada y eficiente en el entorno real de trabajo.

Conclusiones

Gracias a este proceso de Lead Scoring basado en Machine Learning:

- Optimizamos la clasificación y priorización de prospectos.
- Mejoramos la eficiencia de los equipos de ventas y marketing.
- Logramos una **precisión del 88.4%** con Random Forest.
- Diseñamos un flujo de trabajo automatizado para la toma de decisiones basada en datos.

Próximos pasos:

- Seguir monitoreando el rendimiento del modelo.
- Probar otras técnicas de optimización.
- Implementar mejoras en la interpretabilidad del modelo.