

Segmentación de clientes para personalización de campañas de marketing.

Introducción

Este proyecto se basa en un conjunto de datos proveniente de campañas de marketing directo realizadas por una institución bancaria portuguesa. Las campañas se centraron en contactar a los clientes por vía telefónica, y el objetivo principal era predecir si un cliente suscribiría o no un depósito a plazo tras recibir múltiples llamadas.

Se busca desarrollar un modelo de machine learning para agrupar a los clientes en diferentes segmentos basados en sus características y comportamientos, lo que es útil para personalizar las campañas de marketing.

Objetivos del Proyecto

¿Cuáles son los objetivos del negocio?

El objetivo principal del negocio es identificar y clasificar diferentes grupos de clientes según sus características y comportamientos de gasto. Esto permitirá a la empresa personalizar sus ofertas y campañas de marketing, mejorando así la relación con los clientes y aumentando la tasa de retención. Además, la segmentación facilitará la identificación de oportunidades para lanzar nuevos productos adaptados a las necesidades específicas de cada grupo de clientes.

La segmentación puede ayudar a la institución bancaria a dirigir sus esfuerzos de marketing de manera más efectiva, ofreciendo productos específicos que se alineen con las necesidades de cada segmento.

¿Qué decisiones o procesos específicos desean mejorar o automatizar con ML?

Con el uso de machine learning, se busca automatizar el proceso de segmentación de clientes mediante el análisis de características demográficas y financieras. Esto permitirá que la empresa tome decisiones más informadas sobre estrategias de marketing y ventas, optimizando la asignación de recursos y maximizando el retorno de la inversión. Por ejemplo, se pueden personalizar las campañas de marketing dirigidas a diferentes segmentos, basándose en la probabilidad de respuesta de cada grupo.

¿Se podría resolver el problema de manera no automatizada?

Sí, el problema de la segmentación de clientes podría resolverse de manera no automatizada mediante el análisis manual de los datos y la creación de segmentos basados en criterios como la edad, el estado civil o el balance bancario. Sin embargo, este enfoque sería laborioso y propenso a errores, además de no ser escalable. La automatización a través de machine learning permite un análisis más rápido y eficiente, capaz de adaptarse a cambios en los datos y en el comportamiento de los clientes a lo largo del tiempo.

Metodología Propuesta

¿Cuál es el algoritmo de Machine Learning más adecuado para resolver este problema?

Algoritmo: (clustering) como K-means

Justificación:

- Permite clasificar datos en grupos basados en características similares, ayuda a descubrir patrones en los datos que pueden no ser evidentes a simple vista, facilitando la identificación de grupos de clientes con comportamientos y características similares. El algoritmo se basa en agrupar los datos en clústeres al calcular la distancia entre los puntos de datos y los centroides de los clústeres, lo que resulta fácil de entender y aplicar.
- Puede manejar grandes conjuntos de datos y puede ser implementado fácilmente en Python.
- Es más, óptimo a nivel de procesamiento que aplicar una red neuronal.

Métricas de evaluación:

- **Elbow Method:** Utilizado para determinar el número óptimo de clústeres al observar la disminución de la inercia a medida que se incrementa el número de clústeres. El "codo" en la gráfica ayuda a identificar el número ideal de clústeres. Es por ello que **antes** de aplicar K-means para determinar el número óptimo de clústeres que se deben utilizar en el modelo se aplica esta métrica.
- **Silhouette Score:** Esta métrica mide cuán similares son los objetos dentro del mismo clúster en comparación con los objetos en otros clústeres. Un valor más alto indica una mejor separación entre los clústeres.
- **Inercia:** Mide la suma de las distancias al cuadrado entre los puntos y el centroide del clúster al que pertenecen. Un valor más bajo indica que los puntos dentro de un clúster están más cerca del centroide.

Dades Disponibles

El conjunto de datos incluye información detallada sobre las campañas de marketing telefónico realizadas por un banco, incluyendo características demográficas y financieras de los clientes con un total de 41,188 registros y 20 características. Entre las variables disponibles se encuentran la edad, ocupación, estado civil, nivel educativo y tipo de contacto. Además, se incluyen datos bancarios clave, como si el cliente tiene créditos en mora, el saldo medio anual de su cuenta, y si posee préstamos personales o hipotecas.

Métrica de Éxito del Proyecto

La métrica de éxito para este proyecto de segmentación de clientes será el Aumento en la Retención de Clientes. Esta métrica se medirá mediante el porcentaje de clientes que continúan utilizando los servicios del banco después de implementar estrategias de marketing personalizadas basadas en la segmentación

Responsabilidades éticas y sociales

1. Protección de la Privacidad de los Clientes:

- Asegurar el derecho de privacidad de los clientes al recopilar y analizar sus datos.

2. Transparencia en el Uso de Datos:

- Los clientes deben ser informados sobre cómo se recopilan y utilizan sus datos.

3. Evitar Discriminación:

- Al aplicar algoritmos de segmentación, se debe garantizar que no se discrimine a grupos de clientes basándose en características demográficas, socioeconómicas o de comportamiento.

4. Cumplimiento Legal:

- Asegurarse de que todas las prácticas de recopilación y análisis de datos cumplan con las leyes y regulaciones sobre protección de datos como GDPR.

Git: <https://github.com/Nathifer/Project-Machine-Learning/>