Texto, Pizarra

Descripción generada automáticamente

**Licenciatura en Inteligencia artificial y robotica**

**Trabajo practico Nro. 4**

**Seminario de practica en ciencia de datos**

**Segmentación de mercado con Machine Learning**

**Autor: Donadío Gerónimo**

**Fecha de entrega: 17/11/2024**

El objetivo de este informe es encontrar información valiosa para el marketing en nuestra bases de datos de clientes. Para esto trabaje con varios algoritmos del cual el que brindo los mejores resultados fue el aglomerativo. Tuve muchos problemas con el algoritmo k-means, nose si lo implemente mal yo pero los resultados fueron horribles, con el db scan tambien, hasta usando k neighbors para el dbscan, pero no me dio buenos resultados, pero con el aglomerativo tuve buenas predicciones.

Para validar los clústeres generados en el análisis, se utilizaron dos enfoques principales: el método del codo y el cálculo del índice de Silhouette.

**Método del codo**

**Gráfico, Gráfico de líneas

Descripción generada automáticamente**

En el análisis inicial, se empleó el método del codo para determinar el número óptimo de clústeres. Este método evalúa la variabilidad explicada (WCSS, "Within-Cluster Sum of Squares") en función del número de clústeres k. La gráfica resultante mostró cómo WCSS disminuye al aumentar k, indicando una mejora en la compactación de los clústeres. Sin embargo, este descenso se estabiliza en un punto, formando un "codo", lo que sugiere el número óptimo de clústeres. En este caso k=4, lo que indica una buena separación de los datos en grupos significativos.

**Índice de Silhouette**

Posteriormente, se calculó el índice de Silhouette, una métrica que evalúa qué tan bien se agrupan los datos dentro de sus respectivos clústeres y qué tan separados están de los demás clústeres. El puntaje promedio obtenido fue favorable,( 0.2620) lo que indica que los clústeres formados tienen una cohesión interna sólida y una buena separación entre grupos. Este resultado refuerza la validez del número óptimo de clústeres seleccionado previamente.

El uso de estas dos técnicas complementarias asegura que los clústeres son representativos y tienen sentido desde el punto de vista analítico. El puntaje alto de Silhouette y el patrón observado en el método del codo sugieren que los resultados son robustos, lo que respalda la calidad del análisis de clústeres realizado.

Visualizacion de los cluster:

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Para visualizar los clústeres generados,utilize una técnica de reducción de dimensionalidad, específicamente **PCA (Análisis de Componentes Principales)**. Dado que los datos originales tienen múltiples dimensiones (variables), PCA proyectó estos datos en un espacio de dos dimensiones, conservando la mayor cantidad de variabilidad posible. Esta transformación permitió representar gráficamente los clústeres en un plano bidimensional, donde cada punto corresponde a un cliente y los colores indican a qué clúster pertenece. La validación visual confirmó que los clústeres generados no solo tienen sentido desde una perspectiva cuantitativa, sino que también son coherentes desde el punto de vista gráfico. Este tipo de análisis es particularmente útil para detectar problemas, como la superposición significativa entre clústeres (lo que indicaría una separación insuficiente) o la presencia de grupos desiguales en tamaño (lo que podría reflejar desequilibrios en los datos).

De acá podemos sacar varias conclusiones muy valiosas, muestra dos grandes grupos principales, representados por las ramas de color **naranja** y **verde**. Estos grupos se separan a una distancia relativamente grande (alrededor de 120 en la escala del eje Y). La gran separación a una altura alta indica que estos dos grupos son bastante distintos entre sí, lo que sugiere que tus datos se pueden dividir en al menos dos clusters importantes. Dentro de cada uno de los dos grupos principales (naranja y verde), hay varias subramas que se dividen a distancias más bajas (entre 20 y 40). Esto indica que, aunque existen dos clusters principales. Ahora lo importante, a donde llegamos con todo esto en el marketing.

Los grupos principales podrían representar clientes con diferentes niveles de compromiso, comportamiento de compra o capacidad de gasto. A continuación, se describe lo que podrían representar cada uno de los dos principales clusters, el cluster naranja representa clientes de bajo compromiso o con menor poder adquisitivo, estos hacen menos compras, gastan menos o compran productos de bajo costo, pueden ser mas sensibles a descuentos, promociones y programas de fidelización. Por otro lado, el cluster verde, representa clientes de alto compromiso o con mayor poder adquisitivo estos clientes podrían ser los que realizan compras frecuentes, compran productos de alto valor o participan más activamente en campañas de marketing es probable que este grupo tenga un mayor valor de vida del cliente (LTV) y pueda representar a los clientes más rentables. Dentro de los subcluster podría haber variaciones mas especificas, compradores ocasionales vs compradores frecuentes, clientes leales vs clientes nuevos entre otras. Aplicando esto podemos ver que podemos lograr una segmentación precisa en lugar de tratar a todos los clientes de la misma forma, se pueden desarrollar estrategias específicas para cada cluster, por ejemplo ofrecer productos premium o experiencias exclusivas a los clientes del cluster de alto valor y en contra parte enviar descuentos y promociones a los clientes más sensibles al precio del cluster de bajo valor.

Posibles estrategias para el cluster verde:

* Programas de fidelización: Implementar programas exclusivos que ofrezcan recompensas por la lealtad, como descuentos VIP o acceso anticipado a productos.
* Upselling y cross-selling: Dado que estos clientes tienden a gastar más, podrías diseñar estrategias de upselling (productos premium) o cross-selling (productos complementarios).
* Experiencias personalizadas: Ofrecer servicios personalizados, como atención al cliente exclusiva o recomendaciones basadas en el comportamiento de compra.
* Campañas de retención: Desarrollo de estrategias para retener a los clientes clave, como promociones para su cumpleaños o eventos especiales.

Posibles estrategias para el cluster naranja:

* Ofertas promocionales: Campañas centradas en precios más bajos podría aumentar el compromiso.
* Aumentar la frecuencia de compra: Diseño de estrategias para convertir a estos clientes en compradores más frecuentes, como recordatorios de productos o campañas de retargeting.
* Educación del cliente: Enviar contenido informativo o educativo sobre productos que podrían interesarles, aumentando la posibilidad de que se conviertan en compradores más comprometidos.