

Análisis Clúster de Datos en el Sector de Activación de Fibra Óptica para Tigo Business: Estudio de Servicios de Fibra

Maestrante: Raúl Angel Arratia Torrez

Junio, 2024

Introducción

En la actualidad, la activación de servicios de fibra óptica juega un papel crucial en la conectividad empresarial, especialmente en empresas como Tigo Business. Tigo Business se dedica a proporcionar servicios de activación de fibra óptica a clientes ubicados en las ciudades del eje troncal de Bolivia. Estos servicios no solo garantizan una conexión rápida y confiable, sino que también ofrecen diferentes anchos de banda, cada uno con costos asociados que contribuyen significativamente a los ingresos de la empresa.

La implementación de estos servicios no solo mejora la conectividad de los clientes empresariales, sino que también promueve el crecimiento económico al facilitar operaciones comerciales más eficientes y robustas. Este estudio se enfoca en analizar los patrones y características de estos servicios de fibra óptica mediante técnicas de análisis clúster, como clustering jerárquico y K-means, para identificar grupos de servicios que puedan optimizar estrategias comerciales y mejorar la experiencia del cliente.

A través de este análisis, se espera proporcionar recomendaciones valiosas que no solo beneficien a Tigo Business en la optimización de sus servicios, sino también a otros proveedores de servicios de telecomunicaciones que buscan mejorar su oferta en un entorno competitivo y dinámico.

Objetivos

Objetivo General

Evaluar y analizar los patrones de agrupamiento en los servicios de activación de fibra óptica de Tigo Business en las ciudades del eje troncal de Bolivia, utilizando técnicas de análisis clúster.

Objetivos Específicos

1. Identificar los diferentes patrones de comportamiento de los servicios de fibra óptica en función del ancho de banda ofrecido.
2. Aplicar métodos de clustering jerárquico y K-means para agrupar servicios similares basados en características como el costo y el ancho de banda.
3. Analizar la estructura interna de los clústeres identificados para comprender mejor las variaciones en los servicios ofrecidos por Tigo Business.
4. Proponer recomendaciones estratégicas para optimizar la oferta de servicios de fibra óptica y mejorar la satisfacción del cliente.

Motivación

En el contexto dinámico y competitivo de las telecomunicaciones empresariales, la activación de servicios de fibra óptica representa un componente fundamental para la mejora continua de la conectividad y eficiencia operativa de las empresas. Tigo Business, como proveedor líder en el sector, despliega infraestructuras avanzadas de fibra óptica para atender a clientes en las ciudades del eje troncal de Bolivia. La oferta de diferentes anchos de banda adaptables a las necesidades específicas de los clientes no solo optimiza la conectividad, sino que también genera ingresos significativos en función del ancho de banda contratado.

La motivación detrás de este estudio radica en la necesidad de comprender mejor los patrones subyacentes en los servicios de fibra óptica ofrecidos por Tigo Business. Mediante técnicas de análisis clúster, como clustering jerárquico y K-means, este estudio busca identificar grupos de servicios que puedan optimizar estrategias comerciales y mejorar la experiencia del cliente. Estos análisis no solo beneficiarán a Tigo Business en la optimización de sus servicios, sino que también proporcionarán insights valiosos para la industria de las telecomunicaciones en general, promoviendo la innovación y la eficiencia en la oferta de servicios de fibra óptica.

El estudio pretende llenar un vacío de conocimiento al proporcionar recomendaciones estratégicas basadas en datos concretos, lo cual es crucial en un entorno donde la conectividad y la capacidad de adaptación a las necesidades del cliente son imperativos para mantener la competitividad y satisfacción del mercado empresarial.

Marco Teórico

El presente estudio se fundamenta en dos métodos de análisis clúster ampliamente utilizados en la minería de datos y la exploración de patrones en conjuntos de datos:

Clustering Jerárquico

El clustering jerárquico es un método que agrupa los datos en una estructura de árbol o dendrograma. Existen dos enfoques principales: aglomerativo y divisivo. En el método aglomerativo, la distancia entre clústeres se define según la distancia euclidiana:

$$d(C_i, C_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

donde C_i y C_j son dos clústeres, x_{ik} y x_{jk} son las coordenadas de los puntos en los clústeres C_i y C_j , respectivamente, y p es el número de variables.

El algoritmo aglomerativo comienza con n clústeres individuales y en cada paso une los dos clústeres más cercanos hasta que todos los puntos están en un solo clúster.

K-means

El algoritmo K-means particiona n observaciones en K clústeres donde cada observación pertenece al clúster cuyo centroide es más cercano. El objetivo es minimizar la suma de cuadrados dentro de los clústeres (WCSS):

$$WCSS = \sum_{i=1}^K \sum_{x \in C_i} \|x - \mu_i\|^2$$

donde C_i es el conjunto de puntos en el clúster i , μ_i es el centroide del clúster i , y $\|x - \mu_i\|^2$ representa la distancia euclidiana entre el punto x y el centroide μ_i .

El algoritmo K-means converge cuando los centroides ya no cambian significativamente entre iteraciones.

Ambos métodos son útiles para analizar grandes volúmenes de datos y encontrar estructuras no evidentes a simple vista, lo que los hace especialmente adecuados para estudios como el presente, donde se busca identificar grupos de servicios de fibra óptica con características similares para optimizar la oferta y mejorar la satisfacción del cliente.

Descripción de la base de datos

La Base de Datos es proporcionada por la empresa Tigo Business el cual tiene el detalle de todas las activaciones a nivel nacional de los servicios activados en la columna DESCRIPCIÓN contiene los codigos homogeneizados de los servicios como número de cliente, número de cuenta del servicio, nodo, tipo de servicio y anchode banda.

Metodología

La metodología empleada en este estudio se centra en la aplicación de técnicas de análisis clúster para explorar y comprender los patrones de servicios de fibra óptica ofrecidos por Tigo Business. Se optó por utilizar dos métodos principales:

Clustering Jerárquico

El clustering jerárquico se seleccionó debido a su capacidad para revelar estructuras anidadas en los datos, lo cual es ideal para explorar la variabilidad en los servicios de fibra óptica. Este método no requiere la especificación previa del número de clústeres y permite la visualización jerárquica de la relación entre los servicios.

K-means

El algoritmo K-means se eligió por su eficiencia computacional y su capacidad para agrupar eficazmente observaciones en clústeres compactos. Aunque requiere la elección inicial del número de clústeres K , es especialmente útil para identificar grupos de servicios de fibra óptica con características similares en función del ancho de banda y costos asociados.

Ambos métodos se complementan para proporcionar una visión integral de los datos, desde la exploración inicial hasta la identificación y evaluación de patrones de agrupamiento significativos.

Resultados y analisis

Según el siguiente gráfico se puede apreciar que los servicio mas requeridos estan en un intervalo de 30Mbps y 150Mbps, los cuales significan un ingreso medio para la empresa.

Sin embargo, como se observa en la figura existen valores de mas de 200Mbps son esporadicos lo cual significa que las activaciones con ancho de banda mas grandes que el resto, esto denota que los servicios tienen mucho mas costo lo cual beneficia a la empresa.

En el Dendograma se puede apreciar que existe una relación en la activación de servicios puesto que la mayoría de los clientes solo requieren los servicios intermedios porque mas que todo lo usan para conexiones locales, además que la cantidad de usuarios de los clientes muchas veces no sobrepasan los 30 usuarios.

Conclusiones y recomendaciones

De acuerdo a los resultados, se puede concluir que la mayoría de los servicios activados estan en un intervalo intermedio de entre 30Mbps y 150Mbps.

Las activaciones con ancho de banda mayor a 200Mbps son mas escasas lo cual significa menos ingresos para la empresa. No obstante existen un par

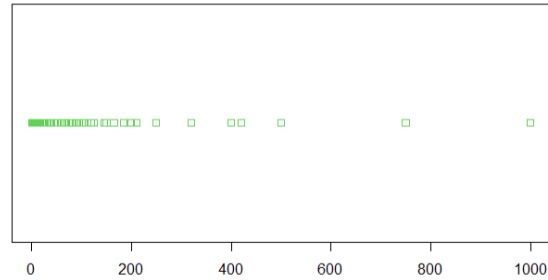


Figure 1: Vizualización del análisis de clustering por el metodo k-medias.

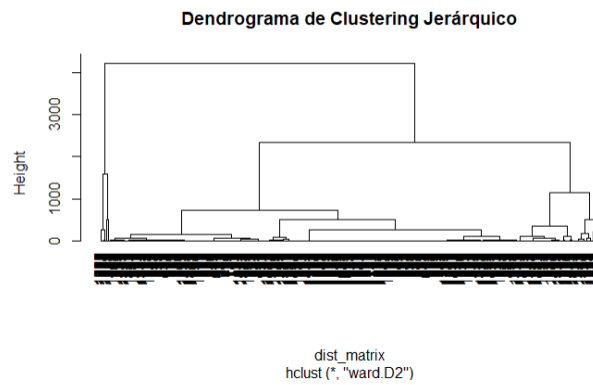


Figure 2: Dendrograma del análisis de clustering jerárquico.

de activaciones de mas de 800Mbps a 1Gbps los cuales tienen un costo muy elevado lo cual beneficia a la empresa.

References

- Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis*. Wiley.
- Jain, A. K. (2010). *Data clustering: A review* (Vol. 31) (No. 3). ACM Computing Surveys (CSUR).
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. , 1, 281–297.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2005). *Introduction to data mining*. Addison-Wesley.

Xu, R., & Wunsch, D. (2015). A comprehensive survey of clustering algorithms.
Annals of Data Science, 2(2), 165–193.

(Xu & Wunsch, 2015) (Jain, 2010) (MacQueen, 1967) (Everitt, Landau, Leese,
& Stahl, 2011) (Tan, Steinbach, & Kumar, 2005)