

Minería de Datos para la caracterización y segmentación de clientes: El caso de la Venta Directa

Julián E. Tornillo^{1,2}, Thomas Gill, Engineer¹, Guadalupe Pascal¹.

¹Facultad de Ingeniería, Universidad Nacional de Lomas de Zamora, Argentina

²Escuela Técnica Superior de Ingeniería Informática, Universidad Rey Juan Carlos, Madrid, España

Abstract— *La industria de la venta directa presenta actualmente elevados niveles de crecimiento en los últimos años, y genera oportunidades de negocio para un gran número de vendedores independientes.*

En estos modelos de negocio, los vendedores tienen objetivos que trascienden la actividad de venta propiamente dicha, como por ejemplo establecer relaciones interpersonales a largo plazo con los clientes y obtener habilidades de administración y gestión de negocios.

En este trabajo se estudia el rendimiento de los vendedores a partir del uso de datos tradicionales de ventas y facturación, combinados con datos de rasgos de personalidad y perfil personal obtenidos a través del test DISC. Se realizan análisis estadísticos y se aplican técnicas de minería de datos; particularmente el análisis de componentes principales y clustering.

Los resultados muestran cuáles son los perfiles deseables para un vendedor que opera bajo esta modalidad y también una clasificación por grupos de clientes combinando estos datos con los tradicionales. Por último, se presentan recomendaciones relacionadas a la ingeniería de ventas en este rubro.

Keywords— *Ingeniería, Minería de Datos, Venta Directa, Personalidad*

I. INTRODUCCIÓN

La industria de la venta directa presenta actualmente altos niveles de crecimiento. Los constantes cambios en la tecnología y en los hábitos de consumo hacen que este rubro sea un objeto de análisis e investigación. El primer autor que estableció una definición formal fue Baker en 1984, definiendo a estos sistemas como una forma de venta sin puntos de venta, distribuidores, mayoristas o cualquier tipo de intermediarios [1]. Según Bennet, el modelo de venta directa es un enfoque de marketing que implica la venta de productos y servicios a consumidores a través de explicaciones personales y demostraciones, principalmente en sus casas [2]. Por su parte, Ostrow y Smith definen a este modelo de negocio desde el cómo se realiza la venta, sosteniendo que estas se deben comúnmente a actividades tales como la venta telefónica, venta puerta a puerta o a través de reuniones (*parties*) en las casas [3]. En la década de los 90 se genera un auge que provoca reinterpretaciones y nuevos conceptos sobre estos sistemas de venta. Varios autores adoptan una visión desde diferentes aristas y estudian en profundidad su impronta operacional, táctica y estratégica. En este sentido, la venta directa es redefinida como un método de distribución de bienes y servicios a través de personal que no trabaja en emplazamientos físicos de una compañía, y lo hace generalmente consolidando pedidos en su domicilio. Desde la

perspectiva operacional, la venta directa es entendida como una forma de comunicación interpersonal o de venta personal. Desde una perspectiva táctica la venta directa puede ser considerada como una forma de organizar las actividades de venta y la función de venta. A nivel estratégico, la venta directa puede ser vista como un canal o modo de distribución, o una manera de penetrar en un mercado para potenciar un negocio [4].

Resulta importante señalar que existen varias posturas que vinculan a la venta directa con los sistemas piramidales. En estos casos, se pone en juicio el componente ético y los potenciales fraudes que ocurren en los sistemas piramidales. Por otra parte, algunos autores resaltan que los sistemas de venta directa contribuyen a la precarización del trabajo y que es usada como método de reducción de costos laborales. Se utiliza como argumento que las personas no trabajan directamente para las compañías bajo relación de dependencia, sino que lo hacen de forma independiente y su subsistencia depende exclusivamente de los márgenes obtenidos por sus ventas personales. Contrariamente, existen posturas que señalan que este modelo permite a un gran número de personas emprender un negocio propio con niveles de inversión significativamente bajos, y brindan la posibilidad de tomar decisiones libremente sobre cómo administrarlo. Adicionalmente, en los sistemas de venta directa, un emprendedor establece sus horarios, su método de distribución, si adopta una estrategia de expansión a partir de redes de vendedores o si se posiciona a partir de altos volúmenes de venta mediante la exposición de sus productos, usualmente en su domicilio o fuera de cualquier establecimiento comercial. En este sentido, se define a la venta directa como un modelo de negocio que presenta oportunidades de desarrollo personal, profesional y económico para todas aquellas personas que deseen obtener ingresos por medio de la generación de un negocio propio [5].

El modelo de venta directa lleva décadas de crecimiento sostenido tanto en popularidad, cantidad de ventas y cantidad de vendedores alrededor del mundo. El rubro presenta sus mayores niveles de crecimiento en épocas de crisis y recesión económica, ya que es una alternativa de generación de ingresos con alto grado de flexibilidad y baja inversión inicial. Actualmente, hay más de 100 millones de personas a nivel mundial que se dedican a comercializar productos y servicios con esta modalidad, ya sea a tiempo parcial o completo. El rubro que más incidencia tiene bajo esta modalidad de venta es el de productos de belleza y cuidado personal, representando más del 50% del mercado global del sector,

como se muestra en la Figura 1. Además este crecimiento se acentúa más en aquellas economías en vía de desarrollo que en los países desarrollados[6].

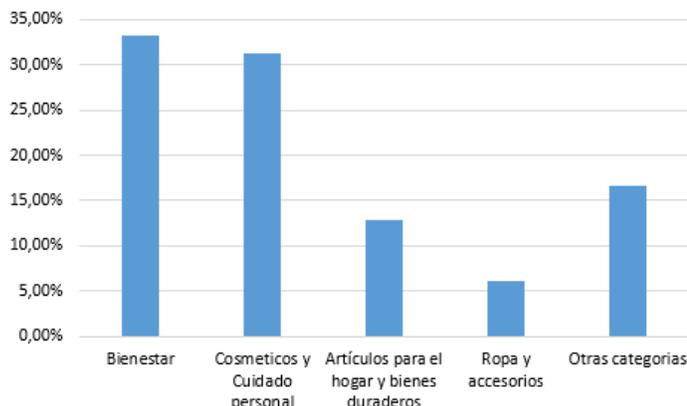


Figura 1: Composición mundial por rubro en 2018. Fuente: Datos WFDSA

La venta directa es considerada hoy en día como un modelo de negocio y es desarrollada tanto por marcas reconocidas a nivel global como por pequeñas empresas locales con perfil emprendedor para comercializar todo tipo de bienes y servicios. Por otra parte, permite a las compañías efectuar la distribución de sus productos y/o servicios a través de vendedores que tienen un comportamiento en conjunto delineado por la organización pero que individualmente poseen independencia total para decidir dónde, cuándo, a quién y cómo realizar las ventas. Este sistema permite a las organizaciones desprenderse de locales físicos donde se tendrían a disposición los productos o servicios que la misma ofrece reduciendo notablemente sus costos fijos. Por otra parte, les brinda a las personas la libertad suficiente para elegir de qué manera llevar a cabo el negocio.

Por otra parte, es notable la incidencia de la personalidad y los perfiles personales en el desempeño y rendimiento dentro de una organización; factor clave para la productividad organizacional [7]. Una de las metodologías más utilizadas en el ámbito empresarial para abordar esta temática es la metodología DISC.

El modelo inicial de la metodología fue propuesto por Marston en 1928 [8] y, en las últimas décadas, se han elaborado una gran cantidad de modificaciones, validaciones y aplicaciones que lo han convertido en una de las técnicas para crear perfiles de personalidad más populares disponibles hoy en día. Desde Clarke en 1956 [9] hasta Watson y Klassen ya en el Siglo XXI [10]. Si bien el modelo presenta un alto grado de simpleza, es una herramienta poderosa que se adapta a las necesidades de las organizaciones si se realiza un adecuado tratamiento y gestión de la información [11]. Existe una forma de mostrar la matriz DISC y es mediante el uso de palabras clave a las que se le asignan a las diferentes letras [12]. Se muestra un esquema de lo mencionado en la Figura 2.



Figura 2: Matriz DISC con palabras clave. Fuente: Everything DISC.

El perfil DISC se define después de responder un cuestionario que es evaluado según cuatro factores que son aquellos que dan nombre al método. Cada una de las letras indica una serie de cualidades y comportamientos en base al estilo de personalidad que representa [13]. Tener baja una letra no es algo negativo, es solo una forma de expresar que esa persona otras características o rasgos diferentes. Las letras son: D (*Dominance/Dominancia*), I (*Influence/Influencia*), S (*Stability/Estabilidad*) y C (*Compliance/Conformidad*).

El perfil se compone del factor con más peso y el de menor peso respectivamente y se expresa con las iniciales de estos factores. Es así como una persona con un perfil DS estaría indicando un gran peso de sus características de dominancia y un bajo peso en estabilidad, dejando a la influencia y conformidad como características intermedias del individuo. Este es el parámetro con el cual se definen los rasgos de personalidad el individuo y que permite definir su perfil personal.

Por otra parte, cuando se cuenta con una gran cantidad de datos, resulta muy complejo realizar análisis que generen conocimiento, ya que estos datos no representan en sí mismos un conocimiento que aumente la competitividad o la productividad en las organizaciones [14]. En el marco del aprendizaje automático, el conjunto de métodos, algoritmos y técnicas que conforman las tareas del aprendizaje no supervisado están orientados a identificar aquella información oculta y de alto valor agregado en una o más bases de datos [15]. Una de las metodologías más utilizadas en este sentido es la Minería de Datos, que está compuesta por diversas técnicas y herramientas para la extracción de conocimiento de una determinada base de datos [16]. En la actualidad, esta metodología se encuentra en permanente evolución. Los grandes volúmenes de datos, la necesidad de procesarlos en tiempo real y la generación de informes y tableros que permitan su visualización ha dado lugar a una nueva organización teórica de estos conocimientos, usualmente llamada ciencia de datos [17]

En este trabajo se realiza una caracterización y segmentación de vendedores que se desempeñan en modelos de negocio con sistemas de venta directa mediante técnicas de análisis cuantitativo de datos. Se estudia la fuerza de ventas, a través técnicas de minería de datos como son el análisis de componentes principales (PCA) y los algoritmos de *clustering* para identificar los grupos que comparten características y comportamientos. Para ello, se analizan datos históricos de una organización que se desarrolla en el rubro de la venta directa en Argentina hace más de 30 años y que es líder en el mercado local. A estos datos se le agregan los resultados del test DISC, el cual fue desarrollado por cada uno de los vendedores.

Con ello se pretende identificar factores clave de éxito y realizar aportes que contribuyan al diseño de estrategias empresariales, por ejemplo, planes de capacitación orientados a cierto grupo específico que necesite desarrollar aptitudes y habilidades claramente determinadas en este trabajo. El análisis toma como base el trabajo de tesis de maestría de uno de los autores del trabajo, en donde definen las variables clave del negocio, se describen estadísticamente y se realiza caracterización preliminar mediante el uso del *software R*. [18].

II. CASO DE ESTUDIO

Este trabajo es una investigación aplicada que se desarrolla a partir de un caso real de estudio. La organización que ha aportado los datos para llevar a cabo esta investigación opera con sistema de venta directa por catálogo y comercializa productos de cosmética, belleza y cuidado personal. Su misión es dar a las personas la posibilidad de desarrollarse personal, profesional y económicamente a partir de un negocio basado en sus productos. Posee además un sistema integrado de gestión en donde se concentra toda la información transaccional, que ha sido una de las principales fuentes de datos utilizada en este trabajo.

La organización tiene un esquema de redes de vendedores, donde a aquellos alcanzan un mínimo volumen de ventas se le asigna un rol de director de ventas (DV). Estos DV se abastecen directamente de la compañía y tienen su propia red de vendedores que de él o ella depende. Si uno de esos vendedores alcanza el volumen de ventas necesario para convertirse en DV, el DV del cual dependía suma una DV generada que se la llamará “Hija”. A su vez, si esa “Hija” genera una DV, se le asigna a la primer DV una “Nieta” y a su vez a la “Hija” se le asigna una “Hija”. Aquella DV que genere “Hijas” y “Nietas” obtendrá descuentos y aumentos en sus comisiones, incorporando a la red una opción estrategia adicional, ya que ahora no solo se puede ganar dinero por la venta en cantidades sino por la generación de hijas y nietas.

Los catálogos se renuevan bimestralmente, a este periodo de dos meses se lo denomina campaña. A cada mes contenido en esos bimestres, se lo denomina “ciclo” y dos de los mismos componen la campaña. Es crucial en la compañía el concepto

de campaña porque se utiliza para la planificación de demanda, diseño de los catálogos, suministros compras y demás procesos administrativos.

Para el mercado y las compañías de la actualidad entender a las personas es una habilidad esencial a la hora de reclutar, gestionar y construir un equipo de trabajo. Tener una mirada comprensiva y objetiva de como las personalidades funcionan es una ventaja importante en el mercado y mejora sustancialmente la competitividad. Es por ello que esta información se vuelve tan relevante en este análisis, puesto que es de gran utilidad a la hora de caracterizar a las personas en el análisis sobre la situación actual. También permite tener una base de datos actualizable para aplicar métodos de predicción y gestionar el futuro de la compañía con el ingreso de nuevos y futuros vendedores.

III. DESARROLLO

La organización almacena información sobre sus DV en sus bases de datos y sistemas de gestión y es ésta la principal fuente de datos de este trabajo. Se analiza el periodo comprendido entre los años 2008 y 2018. Los datos principales a analizar son la edad, el género, los años de antigüedad, cantidad de hijas, nietas, facturación por campaña, facturación anual y el perfil de personalidad DISC. Se realiza una estructuración y un proceso ordenado de análisis, siguiendo la metodología propia de la minería de datos.

Se han considerado los datos de 178 DV, pero se excluyeron siete casos especiales detectados como valores anómalos, sobre todo desde el punto de vista de facturación y generación de descendencia [18].

Los datos incluyen información sobre 171 vendedores y contienen las siguientes variables: 1): ID (Código de identificación de cada vendedor), 2): Edad, 3): Cantidad de años activos en el periodo comprendido entre los años 2008-2017 (Ant.), 4): Cantidad de Hijas DV, 5): Cantidad de Nietas DV, 6): Facturación en la última Campaña (Ult.Cp.), 7): Facturación en el último ejercicio (Ult.Ej.), 8): Perfil DISC o combinación básica (CB). Este último dato se excluye del análisis cuantitativo y se utiliza como indicador para la caracterización en el clustering. Se muestra un ejemplo de los primeros registros en la Tabla 1.

ID	Edad	Ant.	Hijas	Nietas	Ult.Cp.	Ult.Ej.	CB
102	48	10	2	0	222k	1257k	CI
108	45	10	0	0	144k	744k	DC
109	55	10	1	1	206k	660k	CD
113	57	10	3	0	116k	550k	SD

Tabla 1: Ejemplo de estructuración de datos. Fuente: Elaboración propia.

Para realizar el análisis se utilizan técnicas de estadística multivariable y análisis de clúster utilizando los datos de las 171 DV. Las dos técnicas seleccionadas son la de Análisis de componentes principales (PCA) y algoritmos de *clustering*.

Con el análisis de componentes principales (PCA) se pretende generar un nuevo conjunto de variables reducido al máximo que tengan la capacidad de explicar un gran porcentaje de los datos con tan solo unas pocas componentes. Posteriormente se utilizan los resultados de esta técnica para llevar a cabo las tareas de *clustering* agrupando DV.

La unidad observable del PCA es el ID de cliente de cada DV. Las variables consideradas son: 1): Edad, 2): Cantidad de años como DV en el periodo 2008-2017 3): Cantidad de hijas 4): Cantidad de nietas 5): Facturación por campaña (Promedio) 6): Facturación por ejercicio (Promedio).

Para comenzar, habiendo cargado los datos en el software R, se procede realizar el análisis de componentes principales. Este análisis debe incluir una estandarización de los datos para eliminar el problema de la magnitud de los datos y que esto no influya en el resultado del análisis.

En la Tabla 2 se muestran las desviaciones estándar y el porcentaje de representación de cada componente sobre el total.

	PC1	PC2	PC3	PC4	PC5	PC6
Desvío estándar	1.64	1.14	1.02	0.67	0.55	0.45
Proporción	0.45	0.21	0.17	0.075	0.05	0.03
Acumulado	0.45	0.66	0.84	0.91	0.96	1.0

Tabla 2: Incidencia de cada componente. Fuente: Elaboración propia.

La cantidad de componentes principales ideal para analizar este paquete de datos es de tres componentes. La elección se ve sustentada por la Tabla 2 analizada desde distintos aspectos y el criterio de Kaiser. Se puede observar que solo con la primera componente se explica casi el 50% de los datos, el 66% considerando las dos primeras componentes y tomando las tres primeras un 84%. El número de componentes principales a considerar debe ser determinado según su desviación estándar, según el criterio de Kaiser, porque ello representa la variabilidad. Según dicho criterio, se deben considerar aquellas componentes cuya desviación estándar sea mayor a uno. Es por ello que en este estudio se opta por tomar las 3 primeras componentes para continuar el análisis. Esto facilitara, además, la esquematización de los resultados.

Por otra parte, las componentes principales son las combinaciones lineales de las variables originales que explican la varianza de los datos. Estas están compuestas por vectores propios compuestos por un coeficiente para cada variable y se utiliza para indicar la ponderación relativa de cada variable en cada componente. En este sentido se entiende que cuanto mayor sea el valor del módulo, mayor será la incidencia de dicha variable en la componente principal. La tabla 3 permite comprender el criterio de ordenación de los datos en cada componente, considerando aquella variable que tenga mayor valor absoluto en cada una de las componentes. Además, se debe considerar la dirección de cada coeficiente, la cual está dada por el signo.

	PC1	PC2	PC3
1)	0.2231229	-0.70811144	-0.1683899
2)	0.3800878	-0.38502178	-0.4538442
3)	0.4747665	0.01619538	0.4550325
4)	0.3855030	-0.16162674	0.6378364
5)	0.4385254	0.48353040	-0.2038692
6)	0.4893149	0.30024868	-0.3319907

Tabla 3: Criterio de ordenación. Fuente: Elaboración propia.

Si se analiza caso por caso, es en la primera componente donde se evidencia que todas las variables tienen asociaciones positivas. Esto significa que todas ordenan la componente en el mismo sentido. Se podría afirmar que esta componente mide principalmente el rendimiento de ventas y facturación de cada DV. Ya que es esa variable facturación la que posee mayor valor absoluto. Al analizar la segunda componente se observa que el criterio de ordenación está dado en gran medida por la edad (-). Si se considera que la antigüedad (-) también es negativa, se podría afirmar que las asociaciones negativas en esta segunda componente explican cuestiones de temporalidad. Por el contrario, las asociaciones positivas están relacionadas a la venta y la facturación (+).

En el caso de la tercera componente los datos son ordenados según la cantidad de nietas (+), seguido por la cantidad de hijas (+), lo cual puede interpretarse como una componente que mide los niveles de generación de descendencia. Además, cabe señalar que una variable relevante es la antigüedad (+) que desplaza positivamente a la componente.

Las técnicas implicadas para realizar estas tareas son: método de Elbow para definir la cantidad óptima de *cluster*, el algoritmo de K-means para realizar los agrupamientos y el coeficiente de Silhouette y el método de inercias para validar los resultados. Además, se utiliza la combinación básica del *test* DISC para etiquetar a cada DV y favorecer de esta manera la interpretación de los resultados del *test*, combinados con el resto de los datos.

El método Elbow es un método heurístico de interpretación y validación de consistencia dentro del análisis de *cluster* diseñado para ayudar a encontrar el número apropiado en un conjunto de datos. El método Elbow estipula que el número óptimo de *cluster* es aquel en cual adicionando mayor número de los mismos no se reduce significativamente la varianza entre *clusters* [19]. Se observa claramente en la figura 3 que a partir de los cinco *cluster* la curva se vuelve similar a una recta constante, ya que existe un punto de inflexión a partir del quinto cluster.

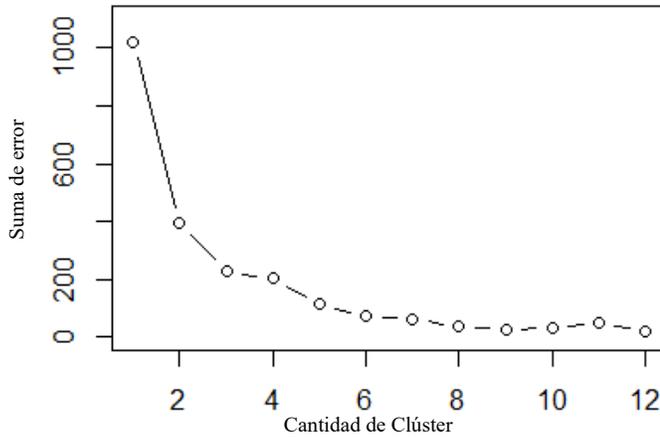


Figura 3: Método Elbow. Fuente: Elaboración propia.

En este sentido, el número óptimo de *cluster* podría ser cinco ya que es el mínimo número con el cual se alcanza un buen desempeño. Sin embargo, también podrían considerarse seis o más y es por ello que resulta importante validar este resultado utilizando el criterio de inercias inter-grupo, como se muestra en la Figura 4, e intra-grupo, en la Figura 5. Luego se muestra un resumen en la Tabla 4.

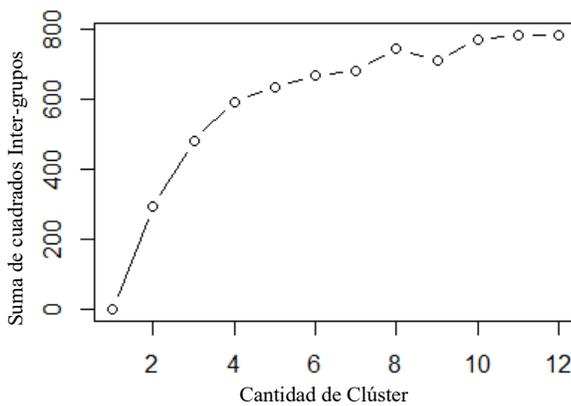


Figura 4: Inercias Inter-grupo. Fuente: Elaboración propia.

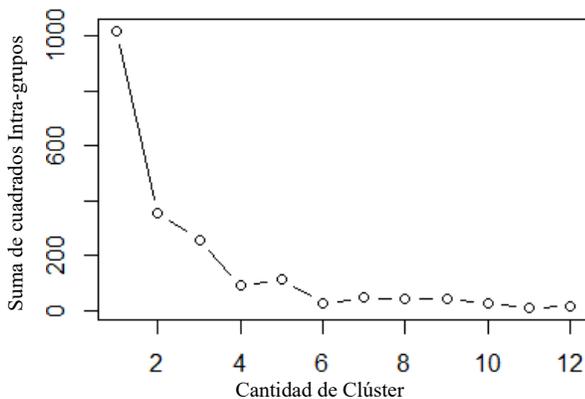


Figura 5: Inercias Intra-grupo. Fuente: Elaboración propia.

Nº Cluster	Silhouette	Inter grupos	Intra grupos
3	0,15	8528779x10 ¹²	1423836x10 ¹²
4	0,07	9043674x10 ¹²	908940507938
5	0,05	9376086x10 ¹²	576528905316
6	0,05	9376086x10 ¹²	576528905316

Tabla 4: Comparación cluster óptimo. Fuente: Elaboración propia.

Se observa que tanto en la evaluación Inter-grupos como en la Intra-grupos se obtendrían muy buenos resultados utilizando cinco *cluster*. Si se utilizaran seis *cluster* se mejoraría levemente el resultado Inter-grupos, pero no habría diferencias sustanciales en la evaluación Intra-grupos y se correría riesgos de incurrir en una clustering excesivo. Por tal motivo, se adopta como valor óptimo para el estudio la utilización de cinco *cluster*.

A continuación, se realiza el análisis de *cluster*, utilizando como datos el resultado del análisis de componentes principales realizado anteriormente. Para ello se utiliza el algoritmo de *k-means*, el cual asigna a cada unidad observable al *cluster* con el centroide más próximo. En la figura 6 se muestra el *clustering* obtenido diferenciándolos por color y en donde la etiqueta de la unidad observable es la combinación básica del test DISC.

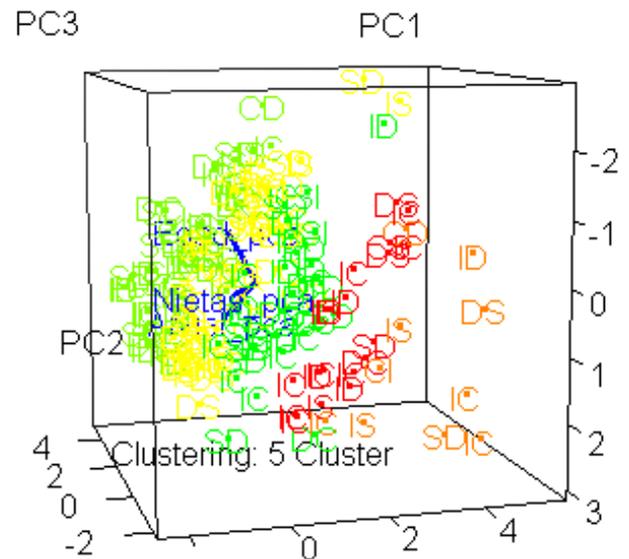


Figura 6: Clustering con PCA. Fuente: Elaboración propia.

Se pueden observar los 5 *cluster* diferenciados por colores y con la etiqueta de la combinación básica. El primer *cluster* presenta altos niveles de venta y facturación, pero niveles de descendencia significativamente bajos. Se interpreta que está compuesto por vendedores autónomos y con alto grado de independencia, ya que esto condice con sus perfiles predominantes, que son DC e IC. Su factor de éxito son su determinación, ambición y confianza.

El segundo está representado por personas jóvenes que están iniciando su carrera en el sector. Los niveles de venta son bajos, así como también la generación de descendencia. Por otra parte, los perfiles DISC en este grupo son diversos y, si bien no son un factor clave que caracteriza al grupo, se interpreta el mayor espectro de acciones posibles para desarrollar el talento de la fuerza de ventas y que lograr que alcancen los niveles de éxito deseados.

El tercer *cluster* contiene aquellos vendedores que tienen una larga trayectoria que sustenta sus logros y son referentes dentro del sistema. En este grupo se encuentran las personas con mayor antigüedad, con una edad por encima del promedio y con niveles de facturación y descendencia notables. De su perfil profesional predominante, con alta dominancia, el principal rasgo característico es la auto-motivación que los impulsa a proponerse nuevos desafíos constantemente.

En el cuarto *cluster* la presencia de jóvenes es nula, ya que contiene a las personas con mayor edad. Los niveles de rendimiento en ventas, facturación y descendencia son moderados y los perfiles predominantes son los SC y CD. Esto permite evidenciar que este *cluster* se presenta poca ambición y pareciera ser que utilizan el negocio como método de distracción y entretenimiento.

El último *cluster* está conformado por personas con altos niveles de ventas, facturación y descendencia. Sin embargo, pese a que el promedio de edad y antigüedad es alto, existen diversos casos con antigüedad menor a 5 años y con altos niveles de descendencia de primer nivel. Los perfiles predominantes son IC e ID. En este sentido, se considera que este grupo está conformado por personas con un marcado perfil empresarial, con resultados sumamente positivos y con gran capacidad de liderazgo y gestión del talento.

Si bien los métodos utilizados han permitido definir que el número óptimo de *cluster* es cinco, resulta necesario evaluar la calidad del *cluster* obtenido. Para ello, se calcula el coeficiente de Silhouette y además se utiliza la valuación de las inercias para concluir que este es el número de *cluster* apropiado para analizar este conjunto de datos. Si bien el mejor coeficiente de Silhouette lo arroja el estudio con tres *cluster*, tanto la inercia inter-grupo como intra-grupo resulta significativamente superior trabajando con 5 *cluster*. Por tal motivo se toma el análisis como válido para continuar con investigaciones futuras.

IV. CONCLUSIONES Y LÍNEAS FUTURAS

En este trabajo presenta una caracterización de la cartera de vendedores del caso de estudio analizado. La integración de los datos tradicionales con los datos de personalidad ha permitido que el análisis sea más completo y se puedan contrastar aspectos cuantitativos con otros cualitativos que hacen al desarrollo personal de los vendedores. Se han identificado aquellos perfiles que más ingresos generan al sistema y cuáles son sus características referidas a su personalidad; como es el caso de los perfiles IC y DS.

Por otra parte, las técnicas de minería de datos utilizadas han proporcionado resultados de interés para la investigación. El *clustering* con 5 agrupamientos ha sido validado y se utiliza como modelo de categorización y segmentación de la fuerza de ventas e insumo para el diseño de planes de desarrollo profesional. Dichos planes deberían ser *ad-hoc* según las necesidades puntuales de cada combinación básica y a qué *cluster* pertenezca. De esta manera la fuerza de ventas es también beneficiaria indirecta de este análisis, ya que la organización podrá poner a disposición capacitaciones, entrenamientos y herramientas conociendo las necesidades puntuales de cada grupo.

Finalmente, una de las líneas futuras de investigación se relaciona con el rendimiento de la fuerza de ventas. Se espera realizar predicciones mediante métodos matemáticos para tener información de cómo sería el rendimiento de un vendedor actual o potencial, partiendo de la base de algunos de los datos utilizados en este trabajo. Además, otra línea en la que se pretende trabajar está orientada a estudiar la permanencia en el sistema de cada vendedor, considerando que en este estudio se ha detectado además un punto de inflexión entre los 2 y 3 años de permanencia. Para ello también se propone utilizar técnicas de predicción, como por ejemplo redes neuronales o el método de los vecinos más cercanos.

AGRADECIMIENTOS

Agradecemos especialmente a la empresa Bagués, que ha suministrado los datos para realizar la investigación. También agradecemos a las instituciones que financian esta línea de investigación, la Universidad Nacional de Lomas de Zamora (UNLZ) a través del instrumento Lomas CyT 2019-2021, al Consejo Interuniversitario Nacional (CIN) a través de las becas EVC-CIN y a la Fundación Atlántica.

BIBLIOGRAFÍA

- [1] M. J. Baker, *Dictionary of Marketing & Advertising*. New York: Nichols Publishing Company, 1984.
- [2] P. D. Bennett, *Dictionary of marketing terms*. American Marketing Association, 1988.
- [3] R. Ostrow and S. R. Smith, *The dictionary of marketing*. Fairchild Publications, 1988.
- [4] R. A. Peterson and T. R. Wotruba, "What is Direct Selling? Definition, Perspectives, and Research Agenda," *Journal of Personal Selling & Sales Management*, vol. 16, no. 4. pp. 1–16, 1996, Accessed: May 09, 2018. [Online]. Available: <http://www.tandfonline.com/doi/abs/10.1080/08853134.1996.10754070>.
- [5] P. DiMaggio and N. W. Biggart, *Charismatic*

- Capitalism: Direct Selling Organizations in America.*, vol. 19. 1990.
- [6] WFDSA, “Global Sales by Product Category - 2019,” Washington D.C., 2019.
- [7] D. Cervone and L. A. Pervin, *Personalidad: Teoría e Investigación 2ª edición*. Mexico: Manual Moderno, 2009.
- [8] W. M. Marston, *Emotions of normal people*. 1928.
- [9] W. V. Clarke, “The construction of an industrial selection personality test,” *J. Psychol.*, vol. 41, no. 2, pp. 379–394, Apr. 1956, doi: 10.1080/00223980.1956.9713011.
- [10] R. Watson and P. Klassen, *Style Insights-DISC: Instrument validation manual*, vol. 28. Target Training International—Performance Systems, Ltd., Success Insights Inc: Phoenix, AZ, 2004.
- [11] J. E. Tornillo, G. Pascal, J. M. Moguerza, and A. Redchuk, “Personality Traits and Business Intelligence: A Model to Improve Direct Selling Systems,” in *2019 5th International Conference on Information Management (ICIM)*, Mar. 2019, pp. 202–206, doi: 10.1109/INFOMAN.2019.8714704.
- [12] A. Bradley, “WORKPLACE PROFILE Your Trusted Source for Everything DiSC,” 2018. Accessed: Mar. 07, 2018. [Online]. Available: www.DiSCProfile.com.
- [13] “Perfil DISC, ‘DISC’.” (Online). Disponible en: <http://www.perfildisc.com>,” 2020.
- [14] L. F. Tabares and J. F. Hernández, “Big Data Analytics : Oportunidades , Retos y Tendencias,” *Univ. San Buenaventura*, p. 20, 2014.
- [15] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to data mining*. 2006.
- [16] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining Practical Machine Learning Tools And Techniques*, no. 9781447150183. 2005.
- [17] L. Cao, “Data science: A comprehensive overview,” *ACM Comput. Surv.*, vol. 50, no. 3, pp. 1–42, Jun. 2017, doi: 10.1145/3076253.
- [18] J. E. Tornillo, “Estudio y Caracterización de Vendedores en Sistemas de Venta Directa en Argentina: Aplicación de Técnicas de Data Mining y Business Intelligence,” Universidad Rey Juan Carlos, 2018.
- [19] P. Sobreiro and D. Martinho, “ANÁLISE DE CLUSTERS PARA SEGMENTAÇÃO DE ESTUDANTES NUMA INSTITUIÇÃO DE ENSINO SUPERIOR,” 2019, [Online]. Available: <http://hdl.handle.net/10400.26/29297>.