

# Análisis del comportamiento de los clientes en las redes sociales mediante técnicas de Minería de Datos

Leidys Contreras Chinchilla, Kevin Rosales Ferreira

leidyscontreras@unicesar.edu.co, kandreyrosales@unicesar.edu.co

Universidad Popular del Cesar, Colombia

**Resumen:** En este documento, se describen los resultados de la utilización de técnicas de minería de datos para analizar el comportamiento de los clientes de una empresa de moda en la red social Instagram. La metodología utilizada fue CRISP-DM a través de la cual se evaluaron los modelos descriptivos utilizando las técnicas de reglas de agrupación y asociación. Los resultados muestran que los modelos propuestos pueden proporcionar información útil para el diseño de estrategias de marketing apropiadas de acuerdo con las preferencias de los usuarios

**Palabras clave:** Minería de datos; Redes Sociales; marketing; clustering; k-means.

**Abstract:** This paper describes the results of using data mining techniques to analyze the behavior of customers of a fashion company in social network Instagram. The methodology used was CRISP-DM through which the descriptive models using the techniques of clustering and association rules were evaluated. The results shows that the proposed models can provide useful information to designing marketing strategies appropriate according to user preferences.

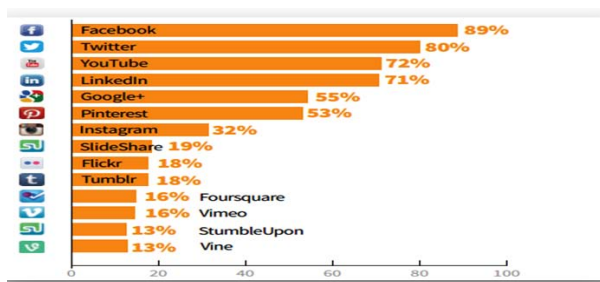
**Keywords:** Data Mining, Social Networks; marketing, clustering, k-means.

## 1 Introducción

Las empresas de hoy tratan de construir sistemas de gestión de relaciones con sus clientes (CRM, por sus siglas en inglés Customer Relationship management), basados en las relaciones sociales del cliente y los patrones de comportamiento [1]. Esto se hace con el fin de elevar el grado de satisfacción de los clientes, lograr retenerlo y fidelizarlo a través de un trato personalizado, adaptando los productos y servicios de la empresa a lo que éste demanda. Las empresas aprenden a entender a sus clientes en el contexto social más amplio y al hacerlo se genera una hipótesis: Las predicciones de conducta de un cliente en el futuro son comportamientos de otros clientes en el pasado. Los patrones de comportamientos de los clientes individuales sirven para construir perfiles analíticos de usuarios. Según este autor, una compañía puede apuntar a un cliente específico con un mensaje específico cuando se tiene un perfil; es decir un perfil personalizado.

Para promocionar los productos a través de diferentes medios las compañías están utilizando las redes sociales disponibles, implementando estrategias B2C (Business-to-Consumer). Éstas consisten en desarrollar estrategias comerciales para llegar directamente al cliente o consumidor final. Las diferentes estrategias son realizadas por el *Consumer Marketer*, quien es la persona que aplicando diversas técnicas y estrategias, apoya el trabajo del equipo de ventas en la comercialización de un producto o servicio [2]. En la Figura 1, se muestra el porcentaje de Marketers B2C que utilizan diferentes plataformas de Social Media para distribuir contenido:

Figura 1. Percentage of B2C Marketers Who Use various Social Media Platforms to Distribute Content



Fuente: [ 2 ]

De la Figura 1, se puede inferir que las empresas están muy interesadas en usar las redes sociales como Facebook, Twitter, Instagram o Youtube, para impulsar sus productos y tener una relación directa con sus clientes. La información que proporcionan las redes sociales permite que las empresas puedan analizar el comportamiento de los clientes a través de diferentes herramientas de mercadeo, algoritmos computacionales, técnicas, modelos estadísticos y matemáticos, con el fin de ajustar sus estrategias empresariales a las necesidades del mercado.

En la actualidad, las compañías están definiendo estrategias para lograr el nivel más alto de lealtad del cliente y se busca la satisfacción del mismo, como paso previo a la fidelización, mediante un marketing individualizado (customized marketing), conversacional y personalizado, con estrategias de retención que involucren al cliente en el proyecto empresarial, incrementen la rentabilidad de la empresa y consigan la lealtad del cliente. Si las empresas no estudian el comportamiento de sus clientes a través de las herramientas tecnológicas como las redes sociales, aprovechando las técnicas de minería de datos para obtener conocimiento valioso de sus clientes, están perdiendo la oportunidad de ser poderosamente competitivas y aumentan el riesgo asociado a la toma de decisiones [3].

Los datos que generan las redes sociales como Instagram proporcionan información muy valiosa a muchas compañías acerca del comportamiento, gustos, preferencias, entre otras características de los usuarios, que es la información que necesitan éstas para desarrollar sus estrategias de mercadeo. Sin embargo, es la transformación de los datos en información y la aplicación de ésta al negocio, lo que generará valor para la empresa [ 4].

Las compañías se encargan de crear bases de datos de gran volumen con información de los clientes para poder aplicar técnicas de exploración y análisis que puedan encontrar patrones significativos de los datos y tomar decisiones acertadas. Es aquí donde se ve la importancia de implementar procesos como la minería de datos o Data Mining en las áreas de publicidad, marketing o cualquier otra en las empresas.

La aplicación de estas técnicas varía de acuerdo al tipo de proyecto a realizar, lo que indica que cada proyecto de minería tendrá ciertas características que harán que se apliquen diferentes técnicas como clasificación, agrupamiento o árboles de decisión. La Figura 2 muestra algunas de las áreas en las cuales se aplican soluciones de minería de datos [5-6]:



Figura 2. Áreas de aplicación Data Mining. Fuente: [5-6]

Para las empresas es de gran importancia conocer las preferencias de sus clientes, para desarrollar estrategias de mercadeo que logren la satisfacción y lealtad de los mismos. Es por ello, que resulta importante aprovechar la información de las redes sociales y las técnicas computacionales modernas para conocer las preferencias y el comportamiento de los clientes de cualquier tipo de empresa.

En este trabajo se realizó un análisis descriptivo del comportamiento de los clientes a una empresa del sector de la moda, de acuerdo al impacto que tiene en las redes sociales, específicamente en la red social Instagram, con el fin de apoyarla en la definición de estrategias de mercadeo relacionadas con los comentarios y likes de sus seguidores en su red social, de manera que se pueda extraer, analizar y proporcionar conocimiento valioso acerca del comportamiento de dichos usuarios.

El resto de este artículo está organizado de la siguiente manera. En la sección 2, se presentan los conceptos teóricos más relevantes y los trabajos previos relacionados con el tema de investigación, en la sección 3, se presentan la metodología utilizada para la consecución de los objetivos. La sección 4 corresponde a la

presentación de los resultados y la discusión de los mismos y, finalmente, en la sección 5, se presenta las conclusiones y los trabajos futuros.

## 2. Antecedentes

### 2.1. Minería de datos (Data Mining)

Es un campo de las ciencias de la computación que consiste en la exploración y análisis por medios automáticos o semiautomáticos, de grandes cantidades de datos con el fin de descubrir patrones significativos [7]. Otra definición de minería de datos es la extracción de interesantes (no-trivial, implícito, previamente desconocido y potencialmente útil) patrones o conocimiento de una enorme cantidad de datos [8]

#### 2.1.1 Tareas de Minería de datos

El proceso de aplicar minería de datos está dado por una serie de modelos (predictivos y descriptivos) y unas tareas para la extracción y uso de la información que están representados en la Figura 3.



Figura 3. Modelos y tareas en minería de datos. Fuente: Elaboración propia basado en [7]

#### 2.1.2 Metodologías utilizadas en minería de datos

Se han desarrollado metodologías cuyo objetivo es la implementación de proyectos de minería de datos, entre las cuales se tienen SEMMA (Sample, Explore, Modify, Model, Access), CRISP-DM (Cross Industry Standard Process For Data Maning) y Microsoft. La Tabla 1 muestra un resumen de las metodologías de procesos de minería de datos teniendo en cuenta las actividades que realizan:

Tabla 1. Metodologías de Minería de Datos y actividades

METODOLOGÍAS	CRISP-DM	SEMMA	MICROSOFT
Estructura	Fases y niveles	Fases	Fases
Niveles	Parte de lo general a lo específico	No tiene	No tiene
Fases	Análisis del problema Análisis de datos Preparación de datos Modelado Evaluación Explotación	Muestreo Exploración Manipulación Modelado Valoración	Definir el problema Preparar los datos Explorar los datos Generar modelos Explorar y validar los modelos Implementar y actualizar los modelos
Herramientas	Genéricas	SAS	Microsoft
Procesos	Iterativo e interactivo entre fases	Iterativo e interactivo entre fases	Iterativo e interactivo entre fases

Documentación	Modelo de referencia Guía de usuario	No se especifica	No se especifica
Objetivos	Se centra en los objetivos empresariales del proyecto	Se centra en las características técnicas del desarrollo del proceso	Se centra en los objetivos empresariales del proyecto

Las tres metodologías están compuestas por etapas o fases que interactúan entre sí. La tecnología SEMMA está más ligada a los aspectos técnicos de la exploración de datos. La CRISP-DM y Microsoft están más enfocadas en el negocio. Esta diferencia se ve específicamente en la primera etapa donde SEMMA comienza con el trabajo de datos mientras que CRISP-DM y Microsoft empiezan por el estudio del negocio y sus objetivos.

### 2.1.3 Análisis realizados en redes sociales utilizando técnicas de minería de datos

La Tabla 2 muestra algunas investigaciones relacionadas con el objeto de esta investigación

Tabla 2. Aplicaciones de minería de datos en redes sociales

TIPO DE EMPRESA	RED SOCIAL UTILIZADA	TECNICA DE MINERÍA	TRABAJO REALIZADO
TIENDA DE ROPA	FACEBOOK	Algoritmo de agrupamiento llamado GAD	<b>Detección de Spam</b> Se analizó la cuenta de la empresa de ropa Hollister.co, y se identificó que 4 de cada 6 fotos recientes son spam y no precisamente hacían alusión a la marca de la compañía o publicidad
PIZZERÍA	FACEBOOK y TWITTER	Análisis de texto no estructurado ó Text mining	<b>Fidelización de clientes</b> Los clientes compartieron su sentimientos y emociones, incluyendo tanto las experiencias positivas y las negativas El análisis arrojó que estas cadenas de pizza participan activamente con sus clientes en las redes sociales y las usaron no sólo para promover sus servicios, sino también para fidelizar a sus clientes
EDUCACIÓN	MySpace	Redes neuronales (Mapas de Khonen)	<b>Análisis de emociones</b> Se determinó que dos tercios de los comentarios expresaban emociones positivas y que tan sólo un 20% contenía emociones negativas. El estudio también pudo determinar que las mujeres son propensas a dar y recibir más comentarios positivos que los hombres.
		Algoritmo K-means	<b>Análisis de deserción</b> Posibles causas de deserción en cada periodo académico, bajo el estudio de variables demográficas del alumno con el registro de última matrícula del semestre de abandono y las causas que lo generaron. Se lograron identificar patrones y tendencias de los alumnos desertores

Fuente: Elaboración propia con base en [10-13]

## 2.2.Red Social Instagram

Se trata de una aplicación móvil y red social para compartir fotografías y vídeos, generalmente creados por los propios usuarios, aprovechando un abanico de herramientas de retoques y filtros fáciles de usar. Mientras en Pinterest y Tumblr se enfatiza el uso de imágenes ajenas (que pueden ser fotografías, pero también gráficos o ilustraciones, así como vídeos), en Instagram se resalta la creación y el compartir fotografías y vídeos creados por el usuario, de móvil a móvil. Mientras Pinterest ordena la información visual para coleccionarla y Tumblr para compartirla, Instagram ha instaurado un formato fotográfico y una interfaz que invitan a señalar con “me gustas” las imágenes preferidas sin detenerse demasiado en cada una de ellas. Pinterest y Tumblr tienen versiones para móvil, pero son redes pensadas mayoritariamente para la comodidad de navegación web o en tableta. Por su parte, es posible consultar el perfil propio de Instagram en la web mediante aplicaciones externas, pero las acciones propias de red social han de realizarse desde el móvil [14].

Instagram es una red social que cuenta con más de 300 millones de usuarios por mes, 70 millones de fotos y 2.5 billones de “me gusta se suben diariamente, se hacen alrededor de 1000 comentarios por segundo. Al existir una red con estas características, las empresas están apuntando a crear marketing basado en social media y no sólo en el marketing tradicional, proporciona una cantidad de información considerable acerca de sus usuarios. Dichos datos pueden extraerse de forma gratuita a través de las API que ponen a disposición las empresas de redes sociales para los desarrolladores de aplicaciones [15].

El hecho de que se pueda extraer información de usuarios de forma gratuita de una red social como Instagram, hace que se pueda analizar una gran cantidad de información de personas que constantemente interactúan con otras y con empresas a través de fotos y comentarios sin tener que realizar encuestas o acceder a base de datos de empresas. Al disponer de esta información valiosa de los usuarios que siguen a cierta empresa a marca en Instagram, se pueden desarrollar técnicas de minería de datos que permitan explorar a profundidad patrones, tendencias y lograr conocimiento útil de las preferencias de los usuarios respecto a productos como ropa o accesorios, como lo es el caso de la empresa en estudio.

Es por esto que en este proyecto se aprovecha la disponibilidad de información de los usuarios de la red social Instagram para realizar estudios de mercado basados en técnicas de minería de datos teniendo en cuenta los datos proporcionados por los mismos usuarios y crear una base de datos propia sin pedir información a otras empresas u organizaciones.

## 3. Metodología

La metodología que se usó para la realización de este proyecto es la metodología llamada CRISP-DM (Cross Industry Standard Process For Data Mining). La metodología CRISP-DM se describe en términos de un modelo de proceso jerárquico, que consiste en un conjunto de tareas que describen cuatro niveles de



abstracción: fase, tarea genérica, tarea especializada y la instancia de proceso [16].

Se seleccionó esta metodología por la independencia de la herramienta tecnológica a utilizar en la exploración de datos, lo que la hace muy flexible, como se muestra en la Tabla 4. Esta metodología consta de las siguientes fases como se muestra en la figura 4 [17].

- **Entendimiento del negocio.** Enfocada a entender los objetivos y requerimientos del proyecto desde la perspectiva del negocio.
- **Entendimiento de los datos.** Comienza con la recolección de datos. En ella, se realizan actividades con el propósito de familiarizarse con los datos.
- **Preparación de Datos.** Cubre todas las actividades necesarias para construir el subconjunto de datos final (dataset) desde los datos originales (sin procesar).
- **Modelado.** Enfocada a la selección y aplicación de la técnica de minería que servirá para obtener un modelo para representar el conocimiento.
- **Evaluación.** Hay que revisar los pasos ejecutados en la construcción del modelo para asegurarse que éste alcanza los objetivos del negocio para la toma de decisiones.
- **Desarrollo.** La implementación de esta fase puede ser tan simple como generar un reporte o tan compleja como implementar un proceso de Minería de Datos repetible a través de la empresa.

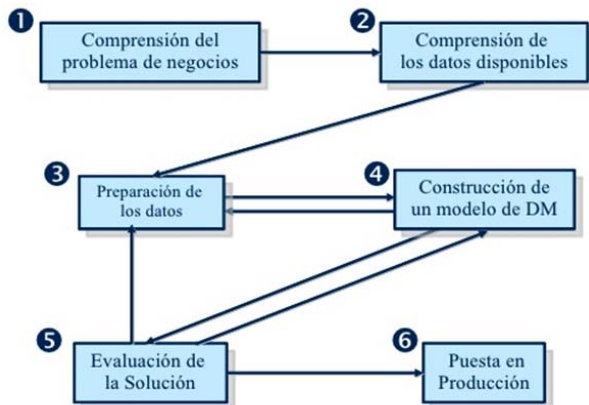


Figura 4. Fases de la metodología CRISP-DM

## 4. Resultados y discusión

Una vez aplicada la metodología CRISP-DM a los datos extraídos de la Red Social Instagram de la empresa en estudio se obtuvieron los siguientes resultados:

### 4.1 Entendimiento del negocio

Empresa de moda líder a nivel nacional que viste a prestigiosas figuras de la farándula, cine y espectáculo e incluso a reinas de belleza. Esta empresa comparte frecuentemente imágenes y videos de sus últimos diseños, accesorios, eventos y tendencias en el ámbito de la moda a través de las redes sociales como Facebook e Instagram.

## 4.2 Entendimiento de los datos

La información procesada corresponde a las imágenes o fotos publicadas por la empresa en estudio en su cuenta de Instagram entre agosto de 2014 hasta abril de 2015. Estos datos fueron extraídos mediante la utilización de la API de Instagram a través de la librería de *python Python-Instagram* para desarrolladores (Ver Tabla 3).

Tabla 3. Datos extraídos de la red social Instagram

Row No	FECHA_CREACION	LINK	TIPO_FILTRO	TIPO_MEDIA	LIKES	COMENTA.	LIKE_USUARIO	TAGS	TIPO_MODAL
75	18/03/2015 18:32	https://www.instagram.com/p/O5-awgmmr/	Normal	image	239	3	False	0	BATA
76	12/03/2015 14:30	https://www.instagram.com/p/O1nZqesZ/	Normal	image	130	0	True	0	BATA
77	13/03/2015 0:59	https://www.instagram.com/p/O3A23qegf/	Normal	image	308	8	True	0	BATA
78	18/03/2015 17:09	https://www.instagram.com/p/O3bA4hul/	Normal	image	271	23	False	0	BATA
79	03/03/2015 17:05	https://www.instagram.com/p/O2tR9Aagf/	Normal	image	188	11	False	0	BATA
80	01/03/2015 23:56	https://www.instagram.com/p/O2tR9Aagf/	Normal	image	222	8	False	0	BATA
81	01/03/2015 9:08	https://www.instagram.com/p/O2p9f1qegf/	Normal	image	233	8	False	0	BATA
82	03/03/2015 22:38	https://www.instagram.com/p/O2p9f1qegf/	Normal	image	347	13	False	0	BATA
83	01/03/2015 13:15	https://www.instagram.com/p/O2p9f1qegf/	Normal	image	370	24	False	0	BATA
84	28/02/2015 20:55	https://www.instagram.com/p/O2p9f1qegf/	Normal	image	258	2	False	0	BATA
85	28/02/2015 19:59	https://www.instagram.com/p/O2p9f1qegf/	Normal	image	380	19	False	0	BATA
86	23/02/2015 18:40	https://www.instagram.com/p/O2p9f1qegf/	Normal	image	412	14	False	0	BATA
87	18/02/2015 23:32	https://www.instagram.com/p/O2p9f1qegf/	Normal	image	177	4	False	0	BATA
88	17/02/2015 15:29	https://www.instagram.com/p/O2p9f1qegf/	Normal	image	207	11	False	0	BATA
89	03/02/2015 18:22	https://www.instagram.com/p/O2p9f1qegf/	Normal	image	386	19	False	0	BATA
90	02/02/2015 12:12	https://www.instagram.com/p/O2p9f1qegf/	Normal	image	206	3	False	0	BATA
91	25/12/2014 15:18	https://www.instagram.com/p/O2p9f1qegf/	Normal	image	221	4	False	0	BATA
92	25/12/2014 11:31	https://www.instagram.com/p/O2p9f1qegf/	Normal	image	254	4	False	0	BATA

### 4.3 Preparación de los datos

El conjunto de datos está compuesto por 1.435 registros y 8 atributos, como lo muestra la Tabla 4.

Tabla 4. Conjunto de datos

NOMBRE CAMPO	DESCRIPCION
FECHA_CREACION	Fecha de publicación
LINK	URL del archivo en Instagram
TIPO_MEDIA	Tipo de archivo (imagen o video)
TIPO_FILTRO	Filtro digital que tiene el archivo
LIKES	Número de "me gusta" de imagen o video
COMENTARIOS	Número de comentarios que tiene la imagen o video
LIKE_USUARIO	Valor booleano que describe si el usuario le dio "me gusta" a su publicación
TAGS	Número de usuarios etiquetados en la imagen o video

Se categorizaron algunos atributos para poder aplicar el algoritmo de clustering, ya que éste sólo permite valores enteros, como se muestra en la Tabla 5:

Tabla 5. Categorización atributos

LIKES	DISCRETIZACION	COMENTARIOS	DISCRETIZACION
0-50	1	0-3C	1
50-100	2	3-6C	2
100-150	3	6-10C	3
150-200	4	>10	4
200-250	5		
>250	6		

TIPO_MODAL	CATEGORIZACION
SOS	1
ACCESORIO	2
BATA	3
BLUSA	4
CAMISA HOMBRE	5
CAMISERO	6
CAMISETA HOMBRE	7
CASUAL	8
COCTEL CORTO	9
COCTEL LARGO	10
FALDAS	11
VESTIDOS DE MATRIMONIO	12

### 4.4 Modelamiento

Las técnicas de modelo descriptivo utilizadas fueron CLUSTERING (Figura 5) con el algoritmo K-MEANS y REGLAS DE ASOCIACIÓN con el operador FP-Growth y Create association rules (Figura 6). Estos operadores se encuentran disponibles en el software RAPIDMINER.

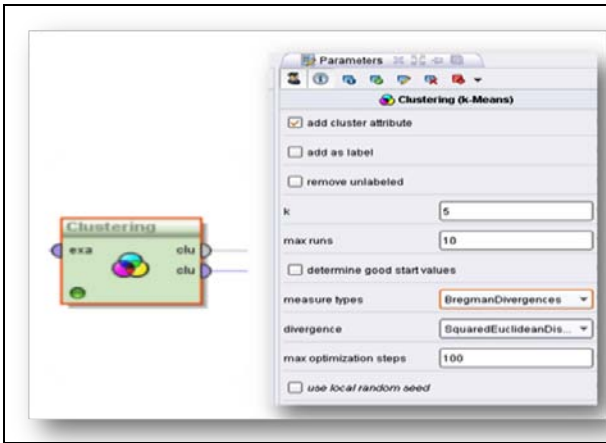


Figura 5. Operador Clustering

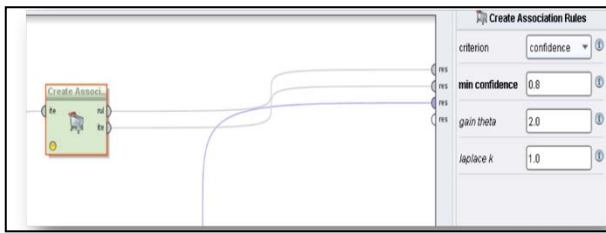


Figura 6. Operador K-means

## 4.5 Evaluación de los modelos

### 4.5.1 Clustering

Una vez aplicado el modelo teniendo en cuenta varios valores de K, se graficó el número de clústeres con el error cuadrático generado por el modelo con el mismo número de K, como se observa en la Figura 7. Analizando los resultados, se observa que en  $k=5$  el SSE es igual a 1,953 y a partir de este valor de k en adelante las variaciones del SSE no son muy significativas.

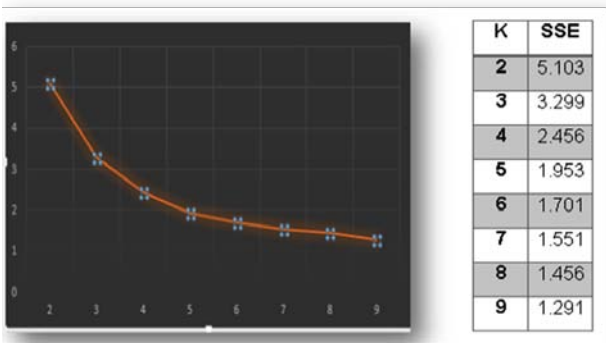


Figura 7. Comparación K Vs SSE

A partir del valor de  $K=5$  se hizo un análisis de los clústeres resultantes ya que no hubo una diferencia en el promedio de valores de los centroides. La descripción del resultado arrojada por software indica que se encontraron 5 clústeres, los cuales contienen la información acerca de las preferencias de los usuarios que siguen a la empresa en estudio. La Figura 8 muestra los clústeres resultantes.

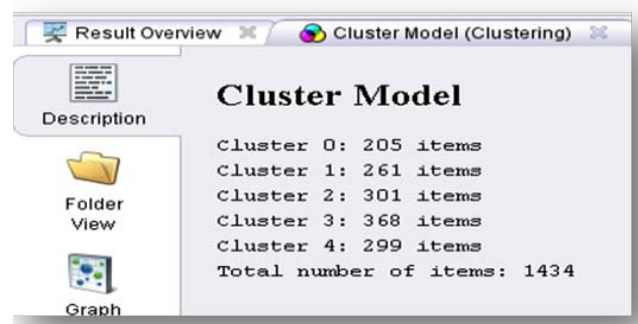


Figura 8. Clústeres resultantes

Tabla 6. Descripción de clústeres

ATRIBUTO	C0	C1	C2	C3	C4
LIKES	200-250	50-100	50-100	50-100	200-250
COMENTARIOS	6-10	0-3	0-3	0-3	6-10
MODA	3	3	8	10	10
TOTAL ITEMS	205	261	301	368	299

### 4.5.2 Reglas de Asociación

Se analizó el conjunto de datos mediante Reglas de Asociación, se configuraron los operadores: **FP-GROWTH** para la formación de los Itemsets frecuentes y **CREATE ASOCIATION RULES** para generar las Reglas de Asociación (Ver Figura 9).

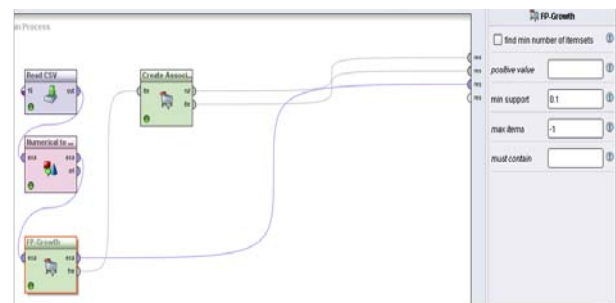


Figura 9. Reglas de asociación

Debido a que el dataset que contiene la información acerca de las publicaciones hechas por la empresa en estudio, no presenta atributos binomiales (1 o 0), debieron ser transformados algunos atributos teniendo en cuenta la fase de entendimiento de los datos y el análisis de clúster. Esto se hace ya que en el operador FP-Growth necesita que todos los atributos de la base de datos de entrada sean binomiales. La Tabla 9 muestra los atributos transformados:

Tabla 7. Atributos transformados

MODAS		COMENTARIOS	
0	COCTEL LARGO, CORTO Y BATAS	0	<=15
1	OTROS TIPOS DE PRENDAS	1	>15

LIKES		FECHAS	
0	<=200	0	2014-2015
1	>200	1	2012-2013

Una vez evaluado el modelo se generaron 9 ítemset frecuentes y 4 reglas de asociación que se muestran en la Figura 10.

No. of Sets: 9	Size	Support	Item 1	Item 2	Item 3
Total Max. Size: 3	1	0.427	TIPO_MODA		
	1	0.404	FECHA_CREACION		
Min. Size: 1	1	0.214	LIKES		
	1	0.043	COMENTARIOS		
Max. Size: 3	2	0.162	TIPO_MODA	FECHA_CREACION	
Contains Item:	2	0.050	TIPO_MODA	LIKES	
	2	0.115	FECHA_CREACION	LIKES	
	2	0.039	LIKES	COMENTARIOS	
Update View	3	0.029	TIPO_MODA	FECHA_CREACION	LIKES

Figura 10. Itemsets frecuentes y reglas de asociación

El modelo fue ejecutado varias veces, con un valor mínimo de soporte de 0.8 (por defecto) desde 0.5 hasta 0.8 (ver tabla 8).

Tabla 8. Confianza vs Reglas generadas

CONFIDENCE	NÚMERO OF RULES GENERATED
0.5	3
0.6	1
0.7	1
0.8	1

Se tomó el valor mínimo de confianza de 0.5, debido a que se generaron más reglas de asociación con respecto a los demás valores en donde solo presentaban 1 regla de asociación. Por lo tanto, las reglas que se generaron al ejecutar el modelo se muestran en la Figura 11.

```

AssociationRules
Association Rules
[LIKES] --> [FECHA_CREACION] (confidence: 0.537)
[TIPO_MODA, LIKES] --> [FECHA_CREACION] (confidence: 0.569)
[COMENTARIOS] --> [LIKES] (confidence: 0.903)

```

Figura 11. Reglas de asociación generadas

## 4.6 Desarrollo

Según los resultados de la evaluación de los modelos se pueden realizar las siguientes conclusiones:

### 4.6.1 Clustering

Teniendo en cuenta la información mostrada en la Tabla 8, se descartan los clústeres 1,2 y 3 porque presentan características similares en sus atributos, se procede a analizar los clústeres 0 y 4 con las siguientes observaciones:

- En el clúster 0 se puede observar que se encuentran 205 fotos donde el atributo más importante es el gran número de “likes” (200-250) y el número de comentarios, que hace referencia a los vestidos tipo BATA. Esto indica que este tipo de vestido es uno de los que más le gustan a los usuarios y por lo tanto hacen muchos comentarios respecto a esta prenda.

- En el clúster 4 se encuentran 299 imágenes en donde se puede observar que el atributo más importante es el tipo de moda, el cual hace referencia a los vestidos de COCTEL LARGO. Cabe resaltar que estos vestidos son los preferidos por los usuarios debido al gran número de comentarios y likes, y es uno de los fuertes de la empresa en estudio, ya que como se puede observar en la Tabla 8 hay 667 publicaciones con este tipo de diseño, agrupados en los clústeres 3 y 4.

### 4.6.2 Reglas de Asociación

Teniendo en cuenta las reglas de asociación generadas que se muestran en la Figura 11, se puede concluir que las fotografías con mayor número de likes y comentarios son las publicadas entre 2012 y 2013.

## 5. Conclusiones y trabajos futuros

Para la empresa del caso de estudio fue de gran importancia conocer las preferencias de sus clientes, ya que esto, es un insumo importante para desarrollar estrategias de mercadeo, basados en el análisis realizado a sus clientes seguidores de la red social Instagram. El análisis permitió establecer que los vestidos preferidos por los clientes son los vestidos de coctel largo, seguido por las batas. También se pudo determinar que los años 2012 y 2013 fueron donde mayor número de likes y comentarios tuvieron estas prendas. Esta información es valiosa puesto que la empresa, puede diseñar e implementar estrategias que tengan en cuenta este tipo de preferencias y logren la satisfacción y lealtad de los clientes impulsando este tipo de prendas.

De igual forma, se pudo evidenciar que el aprovechamiento de la información generada por las redes sociales, en conjunto con técnicas computacionales modernas como la minería de datos, permite a las empresas conocer las preferencias y el comportamiento de sus clientes, sin necesidad de realizar encuestas ni cualquier otro trabajo de campo.

Como trabajo futuro, resultaría interesante probar los modelos de minería de datos desarrollados en esta investigación con otra red social como Facebook y otro tipo de empresas, como las comerciales.

## Referencias bibliográficas

- [1]. TP Hong, LSL Wang IH Ting, *Social Network Mining, Analysis, and Research Trends: Techniques and Applications.*: IGI Global, 2012.
- [2]. AC Martínez, "CRM social: la orientación empresarial hacia el cliente en plataformas 2.0," *REDMARKA. Revista Digital de Marketing Aplicado*, 2011.
- [3]. contentmarketinginstitute.com, "2014 B2C Content Marketing Trends—North America: Content Marketing Institute/MarketingProfs," 2013.
- [4]. Rodríguez Castañeda, "La minería de datos como herramienta de marketing: delimitación y medidas de evaluación del resultado," 2003.

- [5]. Steinbeach, Kumar Tan,., Apr. 2004, ch. 1. [Online]. [http://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap1\\_intro.pdf](http://www-users.cs.umn.edu/~kumar/dmbook/dmslides/chap1_intro.pdf)
- [6]. Micheline Kamber & Jian Pei Jiawei Han, Data Mining: Concepts and Techniques, 3rd ed., 2011.
- [7]. José Hernandez Orallo, Introducción a la minería de datos.: Pearson, 2004.
- [8]. Ethem ALPAYDIN, Introduction to Machine Learning.: The MIT Press, 2010.
- [9]. Ana Azevedo, "KDD, SEMMA AND CRISP-DM: A PARALLEL OVERVIEW ," 2008.
- [10]. Liza Leonor Pinzón Cadena, "Aplicando minería de datos al marketing educativo\*," Notas D Marketing 1, 2011.
- [11]. Choperena, Estrategias de mercadeo a partir de la minería de datos de una empresa del sector de ropa de hogar en Medellín. Envigado: Tesis de pregrado, Escuela de Ingeniería de Antioquia, 2014.
- [12]. Wu He, "Social Media Competitive Analysis and Text mining: A Case Study in the Pizza Industry," Researchgate, 2013. [Online]. [http://www.researchgate.net/profile/Wu\\_He2/publication/257103444\\_Social\\_media\\_competitive\\_analysis\\_and\\_text\\_mining\\_A\\_case\\_study\\_in\\_the\\_pizza\\_industry/links/0046352e2fed3354f5000000.pdf](http://www.researchgate.net/profile/Wu_He2/publication/257103444_Social_media_competitive_analysis_and_text_mining_A_case_study_in_the_pizza_industry/links/0046352e2fed3354f5000000.pdf)
- [13]. David Wilkinson, Sukhvinder Uppal Mike Thelwall, "Data Mining Emotion in Social Network: Gender differences in MySpace," 2009.
- [14]. Raquel Herrera and Luis Codina, "Redes Sociales Visuales: Caracterización, Componentes y posibilidades para," Cuadernos de Documentación Multimedia, vol. 26, pp. 1-24, Sep. 2015.
- [15]. <http://instagram.com>. (2014) [Online]. <http://instagram.com/developer/authentication/>
- [16]. Pete Chapman, CRISP-DM 1.0., 2000.
- [17]. Gregorio Iturbide, "Metodología de Preparación de Datos Orientada a Aplicaciones de," Centro Nacional de Investigación y desarrollo tecnológico, Tesis de Maestría para optar el título de Maestro en ciencias 2013.