

## Segmentación de Mercado Usando Técnicas de Minería de Datos en Redes Sociales

Eduin Olarte<sup>1</sup>, Marisa Panizzi<sup>1</sup>, Rodolfo Bertone<sup>2</sup>

<sup>1</sup> Escuela de Sistemas. Universidad Argentina John F. Kennedy.

Bartolomé Mitre 1411, Ciudad Autónoma de Buenos Aires (C1037ABA), Argentina

<sup>2</sup> Instituto de Investigaciones en Informática LIDI. Facultad de Informática. UNLP – CIC  
eduinolarte@gmail.com; marisapanizzi@outlook.com; rbertone@lidi.unlp.edu.ar

**Resumen.** Las redes sociales han ganado una gran popularidad durante la última década gracias al avance de nuevas tecnologías y al creciente interés de las personas por generar contenidos y compartirlos con sus contactos. Esto hace que los datos generados en las redes sociales crezcan exponencialmente con el tiempo.

Estos datos generados contienen información que se puede ser analizada con el fin de descubrir patrones que ayuden en múltiples disciplinas. El marketing es una de estas disciplinas que está estrechamente ligada a entender comportamientos, tendencias o gustos de las personas

El objetivo de este trabajo consiste en la aplicación de minería de datos (MD) para lograr el descubrimiento de patrones en datos provenientes de redes sociales. A partir de la obtención de patrones se busca realizar distintos tipos de segmentaciones que ayuden a los profesionales de marketing a orientar sus campañas.

**Palabras clave:** Explotación de información, Minería de datos, Segmentación de mercado, Redes Sociales, Aprendizaje Automático.

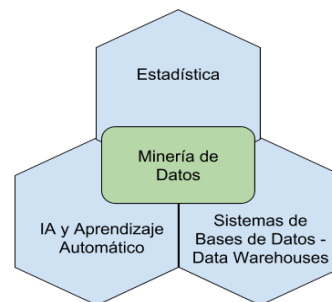
### 1 Introducción

El marketing es un concepto que surge para dar respuesta a la pregunta que se hacen a menudo las organizaciones: ¿a quiénes intentamos vender? El marketing es un proceso social por el cual los grupos e individuos satisfacen sus necesidades al crear e intercambiar bienes y servicios [1]. También es un proceso que comprende la identificación de necesidades, relaciones y deseos del mercado que se tiene como objetivo.

El mercado es demasiado amplio y está formado por clientes diferentes con diversas necesidades. Cada empresa busca centrarse en los individuos, de acuerdo a su capacidad, gusto y segmentos etarios. Con este fin, las empresas realizan estrategias de mercado. La estrategia de dividir el mercado en grupos homogéneos se conoce como segmentación. La segmentación de mercado consiste en dividir un mercado en grupos más pequeños de necesidades, características o comportamientos distintos que podrían requerir productos separados o mezclas de marketing [2].

Tradicionalmente, los datos para realizar las segmentaciones provienen de procesos de recolección tales como encuestas, historial de las ventas, etc. Con el advenimiento de las redes sociales se genera una nueva fuente de información, las redes sociales [3]. Según mediciones internacionales se estima que por minuto se generan más de cuatrocientos mil Tuits los cuales representan una gran cantidad de datos para ser analizados [4]. En la actualidad las redes sociales, proveen privilegios a los desarrolladores para acceder a la información fácilmente por medio de interfaces Web.

Este gran conjunto de datos puede ser analizado para identificar patrones y relaciones. Este análisis puede realizarse con minería de datos. La minería de datos es un campo multidisciplinario que combina estadísticas, aprendizaje automático, inteligencia artificial y tecnología de bases de datos (Figura 1) para establecer relaciones y encontrar patrones en grandes conjuntos de datos [5]. La identificación de patrones y relaciones permite lograr segmentaciones de distintos tipos y dar solución a las empresas.



**Fig. 1.** Disciplinas científicas que comprenden a la Minería de Datos.

La minería de datos es la búsqueda de información nueva, valiosa y no trivial en grandes volúmenes de datos. Es un esfuerzo cooperativo de humanos y computadoras. Los mejores resultados se obtienen al equilibrar el conocimiento de los expertos humanos al describir problemas y objetivos con las capacidades de búsqueda de las computadoras.

En la práctica, los dos objetivos principales de la minería de datos son la predicción y la descripción. La predicción implica el uso de algunas variables o campos en el conjunto de datos para predecir valores desconocidos o futuros de otras variables de interés. La descripción, por otro lado, se centra en encontrar patrones que describen los datos que pueden ser interpretados por los humanos.

Los objetivos de la predicción y la descripción se logran mediante el uso de técnicas de extracción de datos. Basado en el tipo de patrones que se buscan, las técnicas de minería de datos pueden ser clasificación, clusterización, asociación y/o análisis de tendencias [6].

## 2 Análisis de Metodologías para proyectos de MD y herramientas de trabajo

En esta sección, se resume el análisis comparativo de las metodologías para proyectos de minería de datos. Para su análisis se han considerado KDD “Descubrimiento de conocimiento en bases de datos” [7], SEMMA (Sample, Explore, Modify, Model, and Assess [8] y CRISP-DM (Cross Industry Standard Process for Data Mining) [9]. Posteriormente se sintetiza el análisis comparativo de las herramientas de trabajo para los procesos de minería de datos.

### 2.1 Estudio comparativo de las metodologías para proyectos de minería de datos.

El proceso de minería de datos a menudo se caracteriza como un proceso iterativo de múltiples etapas que incluye la selección de datos, la limpieza de datos, la aplicación de algoritmos de minería de datos, la evaluación, entre otros. Se realizó un estudio comparativo de las metodologías con el propósito de identificar si las mismas soportan etapas equivalentes (tabla 1).

**Tabla 1.** Estudio comparativo de metodologías para proyectos de minería de datos.

| Etapas                   | KDD   | SEMMA                                  | CRISP-DM   |
|--------------------------|---|--|--|
| <b>Objetivos</b>         | Pre KDD   |  | Entendimiento de negocio<br>Entendimiento de los datos |
| <b>Pre-procesamiento</b> | Selección<br>Preprocesamiento<br>Transformación | Muestra<br>Exploración<br>Modificación | Preparación de datos                                   |
| <b>Modelo</b>            | Minería   | Modelado                               | Modelado   |
| <b>Resultados</b>        | Interpretación /<br>Evaluación                  | Valoración                             | Evaluación   |
| <b>Despliegue</b>        | Pos KDD   |  | Instalación  |

La metodología KDD cuenta con nueve etapas para el modelo de proceso, CRISP-DM cuenta con seis y SEMMA con cinco etapas. Al examinar los tres modelos de procesos de minería de datos, puede observar que en cierto punto son equivalentes entre sí. Comparándolos, se puede ver:

- El paso del proceso KDD "Desarrollo y comprensión del dominio de la aplicación" se puede identificar con la fase "Comprensión comercial" del proceso CRISP-DM.
- Los pasos del proceso KDD "Creación de un conjunto de datos objetivo" y "Limpieza y pre procesamiento de datos" pueden identificarse con las etapas "Muestra" y "Explorar" de SEMMA, respectivamente, y / o pueden identificarse con la fase "Conocimiento de datos" de CRISP-DM.

- La etapa de proceso KDD "Transformación de datos" se puede identificar con la etapa "Preparación de datos" de CRISP-DM y la etapa "Modificar" del proceso SEMMA, respectivamente.
- Las tres etapas de KDD "Elección de la tarea de minería de datos adecuada", "Elección del algoritmo de minería de datos adecuado" y / o "Algoritmo de utilización de minería de datos" se pueden identificar con la fase de "Modelado" de CRISP-DM y / o "Modelo" etapa del proceso SEMMA respectivamente.
- El paso del proceso KDD "Interpretación de patrones minados" se puede identificar con la fase de "Evaluación" del proceso CRISP-DM y / o la etapa "Evaluación" del proceso SEMMA, respectivamente.
- El paso KDD "Uso del conocimiento descubierto" se puede identificar con la fase de "Despliegue" del proceso CRISP-DM.

En este trabajo, se ha seleccionado como metodología de trabajo KDD ya que en la actualidad es la metodología más utilizada por expertos, además de que se presenta como el campo relacionado al Data Mining. Esto se debe a que es considerado como más completo y preciso, al proporcionar pasos especialmente dedicados a la interpretación y evaluación de los resultados obtenidos con posterioridad a la minería. Por el contrario, CRISP-DM y SEMMA no implementan un estadio de interpretación que para el propósito de este trabajo es muy importante ya que busca detectar patrones de relación útiles, entre usuarios e interpretar qué impacto puede alcanzar una campaña de marketing orientada a esos usuarios

## 2.2 Estudio comparativo de herramientas de trabajo

En este análisis, las herramientas se agruparon en Software Licenciado y Software Libre (Tabla 2).

**Tabla 2.** Herramientas para minería de datos.

| Herramientas de Minería de Datos | Software Libre      | Weka                                       |
|----------------------------------|---------------------|--|
|                                  |                     | RapidMiner                                 |
|                                  |                     | KMine                                      |
|                                  |                     | R  |
|                                  |                     | Python y sus librerías de Minería de Datos |
|                                  | Software Licenciado | SAS Data Mining                            |
|                                  |                     | SPSS Modeler                               |
|                                  |                     | Oracle Data Mining                         |
|                                  |                     | Microsoft SQL Management Studio            |

Para el estudio comparativo (Tabla 3), se consideraron las características más importantes de cada herramienta con el propósito de identificar si las mismas soportan características equivalentes.

**Tabla 1.** Estudio comparativo de metodologías para minería de datos

| Características              | Weka | RapidMiner | R  | Python | Kmine | SQL |
|------------------------------|------|------------|----|--------|-------|-----|
| Clusterización               | Si   | Si         | Si | Si     | Si    | Si  |
| Asociación mining            | Si   | Si         | Si | Si     | Si    | Si  |
| Regresión lineal             | Si   | Si         | Si | Si     | Si    | No  |
| Regresión lógica             | Si   | Si         | Si | Si     | Si    | No  |
| Clasificadores Bayesianos    | Si   | Si         | Si | Si     | Si    | No  |
| Árboles de decisión          | Si   | Si         | Si | Si     | Si    | No  |
| Análisis de series de tiempo | Si   | No         | Si | Si     | Si    | No  |
| Análisis de texto            | Si   | Si         | Si | Si     | Si    | Si  |
| Procesamiento de BigData     | Si   | No         | Si | Si     | No    | No  |
| Flujos Virtuales             | Si   | Si         | No | No     | Si    | No  |

Para el desarrollo de este trabajo se usará Python como lenguaje para aprovechar su dinamismo, su gran variedad de librerías disponibles relacionadas con los campos del aprendizaje automático y que están disponibles como código abierto. También es utilizado ampliamente en ciencias e ingeniería y dispone el soporte de la comunidad que ha estado trabajando en proyectos de Minería de Datos para despejar dudas acerca del desarrollo. Para el propósito de este trabajo Python dispone de la librería Pytweet<sup>1</sup>, que permite el acceso a la Api<sup>2</sup> de Twitter y disponer de los múltiples puntos de acceso que la red social ha dispuesto para los desarrolladores.

### 3. Desarrollo del Caso de Estudio

Con el fin de garantizar una adecuada adquisición de conocimientos, se han realizado una serie de entrevistas con expertos en marketing para tener un dominio general y otro específico, de los aspectos que rodean a la problemática. Las sesiones con los expertos se han dividido de tal forma que ellos participen en pasos claves del desarrollo de este trabajo. Por esto se han desarrollado tres ciclos de entrevistas como parte de las fases de planteamiento de problema, revisión y selección de datos y validación de la solución. Para el relevamiento de información basado en entrevistas se ha utilizado el modelo propuesto por García Martínez R. et al. [11].

La técnica utilizada fue la de Entrevista no Estructurada. De los resultados recolectados surgen las siguientes situaciones:

1. Un problema que puede enfrentar una compañía es que al recolectar datos por medio de encuestas, entrevistas u observaciones, se obtienen datos limitados y estructurados que proveen una visión limitada del mercado. Por otra parte teniendo en cuenta el dinamismo del mercado, estos datos pueden quedar desactualizados rápidamente provocando que la campaña lanzada con posterioridad sea menos acertada.
2. Al generar, en una nueva estrategia de marketing digital por medio de las redes sociales es posible encontrar usuarios que no interesan directamente, pero representan una gran cantidad de casos de análisis. Ante estos casos, la asertividad de la campaña se reduce considerablemente. Estos usuarios pueden ser:
  - Usuarios tipo Bots o de creación automática.
  - Usuarios con poca actividad.
  - Usuarios que viven en un área o país a la que no apunta la campaña.
  - Usuarios con pocos seguidores.
3. Al presentarse un universo de usuarios de interés es necesario analizar cuáles de ellos generan más impacto y entender esas características que los relacionan con otros seguidores. Esto requiere analizar grandes volúmenes de datos e insume

---

1 "Pytweet: Una biblioteca pythonic que proporciona una interfaz simple para la API de Twitter. Los valores se normalizan en tipos de Python.

2 "Application Programming Interface". En la programación de computadoras, una interfaz de programación de aplicaciones es un conjunto de definiciones de subrutinas, protocolos y herramientas para crear aplicaciones de software.

gran cantidad de tiempo que puede producir desactualización del plan de marketing.

En este trabajo se busca dar respuesta a las siguientes preguntas de investigación:

- ¿Es posible obtener datos de redes sociales en tiempo real?
- ¿Es posible filtrar el contenido de aquellos usuarios que no interesan directamente a la campaña de Marketing?
- ¿Es posible realizar segmentaciones de mercado de los seguidores de la cuenta en estudio usando metodologías y técnicas de Minería de Datos?

De acuerdo al objetivo planteado, no se llevará a cabo la instancia de implementación, es decir que no se contempla la instalación del desarrollo realizado. Sin embargo, se realiza el desarrollo conceptual de los modelos paso a paso y se muestran los resultados mediante gráficos y figuras que le permitan al profesional de marketing medir su utilidad y aplicabilidad.

El análisis realizado está limitado a una sola red social, ya que debido a la complejidad de la conexión a las distintas Apis y a las limitaciones de tiempo para realizar consultas, los tiempos del proceso de obtención de datos serían excesivos. Por esta razón también se contempla analizar una sola cuenta ya que obtener datos de otras cuentas necesitaría una cantidad de recursos que no están disponibles. En caso de contar con dichos recursos, el modelo podría escalarse.

Finalmente, el propósito del ejemplo es obtener los Tweets publicados en el ámbito del Capital Federal donde la variable “CapitalFederal” contiene un array con las coordenadas que encierran a la ciudad.

Como entorno de desarrollo se utilizó Anaconda<sup>3</sup> aprovechando sus ventajas para el manejo de paquetes y librerías. Además brinda un tablero de control que permite manejar e instalar los paquetes y aplicaciones para trabajar con el análisis de datos.

Se eligió a Twitter debido a que ofrece ciertas ventajas que la hacen muy interesante desde el punto de vista de análisis de datos. Este trabajo se centra en el análisis de una cuenta con gran número de seguidores de esta red social (Cuenta en Estudio).

La aplicación utilizada para extracción de datos es API REST. Ésta proporciona acceso de lectura sobre los datos públicos en Twitter. No requiere disponer datos de ingreso de las cuentas a las que se pretende acceder, por lo que se pueden extraer datos como los tuits de un usuario, información de perfiles usuario y datos de los seguidores entre otras opciones de una forma muy sencilla. Después un primer análisis en Twitter, se ha detectado como los perfiles de usuario en la red social no contienen información relevante como la edad, el sexo, características físicas. El acceso a la localización de los usuarios es limitado ya que pueden existir tuits localizados o no.

Por lo tanto, el análisis en Twitter enfoca el punto de vista de la actividad de los seguidores, la calidad de estos y sus intereses, datos existentes para cada usuario y accesibles públicamente mediante la API.

---

<sup>3</sup> Es una distribución de código abierto de los lenguajes de programación Python y R para procesamiento de datos a gran escala, análisis predictivo y computación científica, que tiene como objetivo simplificar la administración e implementación de paquetes.

Un factor a tener en cuenta es el límite en la tasa de peticiones. La tasa límite de la API se define en intervalos (o ventanas) de 15 minutos en las que según el método utilizado se pueden realizar un número determinado de peticiones.

En la fase de selección, inicialmente se analizan los datos disponibles relacionados a la cuenta de Twitter seleccionada. Posteriormente obtendrán los datos desde el Api de Twitter seleccionando aquellos datos que se consideren relevantes para el objetivo de este trabajo.

Se genera un Script en Python el cual incluirá los pasos para descargar los datos de estudio de la cuenta seleccionada. Estos pasos son: 1) autenticación y conexión con la Api de Twitter; 2) la obtención de datos de perfil de cada usuario mediante el método GET users/lookup; 3) la descarga y almacenamiento de los últimos 200 tuits de cada uno de los followers, a partir de los IDs obtenidos en la función 2.

Se tomó para el estudio una cuenta de una automotriz ampliamente conocida en Argentina que se unió a Twitter en Octubre del 2011, desde entonces ha publicado más de cuatro mil tuits y llega a más de cincuenta mil seguidores. Haciendo uso de la Api se obtienen datos más detallados. El acceso a estos datos depende de la obtención previa de las credenciales descritas anteriormente para lograr una autorización del tipo OAuth<sup>8</sup>.

En la fase de minería de datos, se plantean escenarios asociados a cada tipo de segmentación (Segmentación Geográfica, Segmentación Demográfica, Segmentación Socioeconómica y Segmentación Conductual), los cuales buscan resolver la problemática desde el punto de vista de cada uno de ellos, analizando los atributos que las caracterizan y aplicando los modelos desarrollados con los datos obtenidos de los estadíos previos. Esto permite comprobar si el escenario de segmentación de mercado es válido y es posible realizar segmentaciones a partir de los datos obtenidos de Twitter.

En este artículo, se presenta el Escenario Segmentación Geográfica. El mismo consiste en realizar una división por países, regiones, ciudades, pueblos o barrios de la información analizada. Twitter permite acceder a la locación del usuario, pero en algunas situaciones con inconvenientes de valores incorrectos o inexistentes.

Para dar solución a este problema previamente se implementó un algoritmo para inferir la ubicación de usuario. Aún en estos casos, existen valores incorrectos que no pueden ser interpretados como una locación. Para ello, se implementa una de las tareas de minería de datos, la clasificación basada en reglas, la cual busca empareja o asociar datos a grupos predefinidos (aprendizaje supervisado).

Estos datos predefinidos son el conjunto de entrenamiento, el cual es preparado por el humano (Experiencia). La implementación del modelo requiere la construcción del conjunto de datos de entrenamiento. Para este fin se ha descargado la base datos del Correo Argentino. Una vez generado el conjunto de entrenamiento, el clasificador lo utiliza como se presenta en la Figura 3, donde determina según la regla que el valor "cap. fed" se refiere a la *Capital Federal, Buenos Aires, Argentina*.

El clasificador basado en reglas hace uso de un conjunto de reglas IF-THEN para la clasificación. No es necesario generar en primera instancia un árbol de decisión. En este algoritmo, cada regla para una clase dada cubre muchas de las tuplas de esa clase. Según la estrategia general, las reglas se aprenden de a una por vez. Para cada vez que se aprenden las reglas, se elimina una tupla cubierta por la regla y el proceso continúa

para el resto de las tuplas. El algoritmo representado en la Figura 4, corresponde a la creación de regla basándose en el conjunto de entrenamiento.

```

Algoritmo: Cubrimiento secuencial
Input: data #Datos de entrenamiento
Output: Un conjunto de reglas IF-THEN.
Method:
rule_set={}; // el listado de reglas iniciales es vacío
for clase in data
    rule = Learn_Rule(clase[location], clase[time_zone] clase[classification_location]);
#Remueve reglas repetidas generadas por entrenamiento previo
Remove_Rules_Covered(rule_set, rule)
# Agrega una nueva regla
Rule_set.append(rule)
return rule_set;
    
```

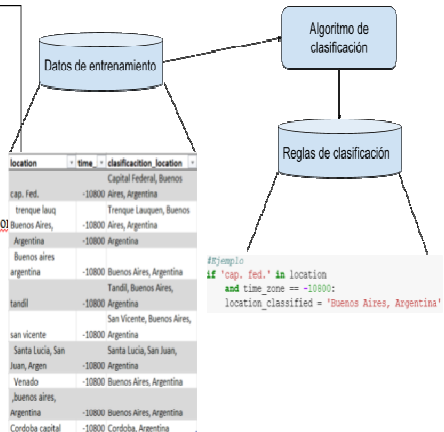


Fig. 3. Algoritmo de generación de reglas

Fig.4. Clasificador de locaciones

Fase de evaluación del modelo: si bien el tratamiento realizado puede brindar un acercamiento a la locación del usuario, pueden existir casos en los que la locación no se obtenga por medio del clasificador implementado. Para estos casos, el campo se deja vacío y no será contemplado para el escenario de segmentación geográfica. La Figura 5 muestra los resultados de la ejecución del modelo. Se obtuvieron que al menos el 70% de los datos fueron clasificados exitosamente mientras el 30% no podrán ser tenidos en cuenta para realizar la segmentación geográfica.



Fig. 5. Resultados de la ejecución del modelo de clasificación

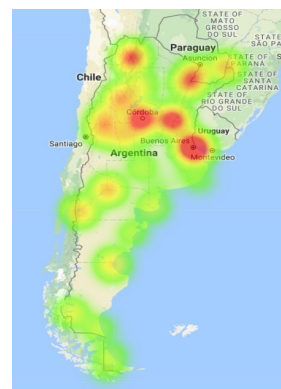


Fig. 6. Concentración de seguidores



Finalmente con las locaciones clasificadas se consiguió detectar la ubicación del 70% de los usuarios de interés que siguen a la cuenta. Con esta información se generó la segmentación geográfica de los seguidores y permitió crear un mapa de calor, el cual indica la densidad de concentración de los seguidores en regiones determinadas (Figura 6).

Otra segmentación geográfica que permite generar la obtención de estos datos fue la división por región, donde se observó que el 80% de los seguidores de importancia de la cuenta están ubicados en la región de la Pampa (Figura 7). Por otra parte, y como lo muestra la figura 8, es posible realizar segmentaciones por provincia.

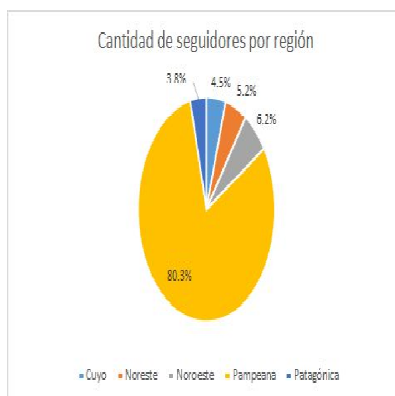


Fig 7. Cantidad de seguidores por Región.



Fig. 8. Seguidores por provincia.

## 4 Conclusiones y futuros trabajos

En este proyecto se ha marcado como objetivo fundamental lograr segmentaciones de mercado a partir de datos obtenidos de redes sociales mediante la utilización de metodologías y técnicas de Minería de Datos.

La elección de Twitter como fuente de los datos resultó satisfactoria ya que provee una potente Api que permite consumir los datos, aunque existieron desventajas a nivel de datos, debido a que Twitter comparte una cantidad limitada de datos de usuarios. Esto no permitió realizar segmentaciones con algunos atributos como por ejemplo el Edad, Género u Ocupación.

Fue necesario elegir un entorno de desarrollo para conectarse a la interfaz expuesta por la red social (Api), se eligió Python para este propósito, lo cual resultó sumamente satisfactorio ya que este lenguaje en la actualidad tiene mucho soporte de la comunidad y cuenta con gran variedad de librerías de código abierto para múltiples propósitos.

Una de estas librerías resultó ser precisa para abstraer el desarrollo de la compleja implementación de consumir una Api y tener que implementar una paginación propia.

La posibilidad de consumir los datos directamente desde los servidores de Twitter brindó respuesta a la primera pregunta problema.

Se aplicó el proceso metodológico propuesto por KDD, por medio de sus procesos permitió solucionar los problemas encontrados. El pre procesamiento generó nuevos datos en base a los obtenidos y utilizarlos como atributos para realizar la segmentación de mercado.

Como trabajos: a) explotar más técnicas de NLP (Neuro Linguistic Programming), por ejemplo investigando sobre el análisis de sentimientos; b) crear una aplicación que permita realizar analytics sobre cualquier cuenta y ofrecer los resultados de la analítica en la misma interfaz.

## Referencias

1. Kotler, Philip; Gary Armstrong, John Saunders, Veronica Wong (2002). «Capítulo 1: ¿Qué es Marketing?». Principles of Marketing (3ª edición europea edición). Essex (Inglaterra): Prentice Hall. ISBN 0-273-64662-1.
2. Lamb, C. W., Hair, J. F., & McDaniel, C. D. (2004). Marketing. South-Western.  
May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148--1158. Springer, Heidelberg (2006)
3. Chen, Z., Kalashnikov, D. V., & Mehrotra, S. (2009). Exploiting context analysis for combining multiple entity resolution systems. In Proceedings of the 35th SIGMOD international conference on Management of data - SIGMOD '09. <https://doi.org/10.1145/1559845.1559869>
4. Digital in 2017: Global Overview - We Are Social. (n.d.). Retrieved October 4, 2017, from <https://wearesocial.com/special-reports/digital-in-2017-global-overview>
5. Ashton Acton, PhD. General Editor. Issues in General Science and Scientific Theory and Method: 2011 Edition. ISBN: 978-1-4649-6346-9.
7. Chen, M.-S., Han, J., & Yu, P. S. Watson IBM Research (1996). Data Mining: an Overview from Database Perspective. <http://hanj.cs.illinois.edu/pdf/survey97.pdf>
8. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. 1996. From data mining to knowledge discovery in databases. AI Magazine, 17(3): 37-54.
9. SAS Institute Inc. 1998. SAS Institute White Paper, From Data to Business Advantage: Data Mining, The SEMMA Methodology and the SAS® System, Cary, NC: SAS Institute Inc.
10. Chapman, P., Clinton, J., Keber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R. 2000. CRISP-DM 1.0 Step by step BI guide.
11. García Martínez, R., & Britos, P. V. (2004). Ingeniería de Sistemas Expertos. Buenos Aires.