

Las Herramientas Informáticas como Estrategia en la Toma de Decisiones Enfocadas en los Impactos Ambientales

Fernando Vázquez Torres¹, María Elena Tavera Cortés¹, Adolfo Bravo Hernández¹

¹ Instituto Politécnico Nacional, Sepi-Upiicsa, Calle Thé 950, Colonia Granjas México, Ciudad de México, D.F., 08400. México

fvazquez@ipn.mx, fvazquez@gmail.com, mtavera@ipn.mx, ofloda@gmail.com

Resumen: El **objetivo** de este trabajo es abordar y facilitar la toma de decisiones desde una perspectiva complementaria apoyada en el uso de herramientas informáticas, sobre el impacto ambiental del uso de los residuos orgánicos [1]. **Metodología:** Esta investigación corresponde a un estudio de caso, diseñado para empresas restauranteras, el método utilizado fue de campo sobre una recolección de datos cuantitativo en 30 restaurantes, utilizando la metodología CRISP-DM (Cross Industry Standard Process for Data Mining) [2]. **Resultados:** Se aplicó a la base de datos obtenida para obtener conocimiento y así lograr una mejor comprensión y análisis de los resultados [3]. **Conclusiones:** El uso de las herramientas de minería en el análisis de la base de datos permitió analizar con profundidad los resultados arrojados. **Discusión:** Las herramientas informáticas nos permiten contrastar entre un análisis tradicional de datos y otro con el uso de las tecnologías dando conocimientos aplicables en forma sistémica y general.

Palabras claves: residuos orgánicos, empresas restauranteras, herramientas informáticas, toma de decisiones, impactos ambientales.

1 Introducción

La presente investigación es el reporte de una investigación para obtener el grado de maestría y dónde se usan herramientas informática, algunas de ellas de minería de datos con la implementación de la metodología Cross Industry Standard Process for Data Mining (CRISP-DM) [2], mediante la herramienta informática RapidMiner, 2015 [5] que permitió trabajar una base de datos obtenida con la aplicación de un cuestionario aplicado en la zona restaurantera de Acapulco para conocer el potencial del aprovechamiento de los residuos orgánicos.

En el mundo de la minería de datos es indispensable contar con un software especializado que permita trabajar los datos para alcanzar los objetivos propuestos. RapidMiner [6] es una herramienta informática de minería de datos, que permite el desarrollo de procesos de análisis de datos mediante el encadenamiento de procesos a través de un entorno gráfico.

La utilización de este software permitirá analizar los datos recabados durante el trabajo de campo, permitiendo no sólo modificar datos, sino también, seguir alimentando la base.

2 Estado del Arte

En 2015, se señala que la utilización de la minería de datos en las actividades cotidianas de las organizaciones en general y de las empresas en particular, se ha convertido en algo cotidiano y más frecuente [7]. Los negocios de distribución y publicidad dirigida han sido las áreas en las que tradicionalmente se han empleado los métodos de minería, ya que han permitido reducir costes o aumentar la receptividad de las ofertas. Pero estas no son las únicas áreas a las que se puede aplicar, de hecho, podemos encontrar ejemplos en todo tipo de aplicaciones: financieras, seguros, científicas (medicina, farmacia, astronomía, psicología, etc.), políticas económicas, sanitarias o demográficas, educación, policiales, procesos industriales, residuos orgánicos entre otros [8].

De acuerdo a [9] la minería de datos es el proceso de: seleccionar, explorar, modificar, modelizar y valorar grandes cantidades de datos con el objetivo de descubrir conocimiento. Asimismo, en [10] se describe como el despertar del interés de los tomadores de decisiones directivas que permitan descubrir en los datos

históricos información útil para fundamentar mejor sus decisiones con la ayuda de estas herramientas. Para nuestro caso en particular, el procesamiento de los residuos orgánicos, se utilizará un sistema diseñado a la medida que permite: 1) agrupar la información en Excel, y 2) graficar los datos y obtener una gráfica de barras.

Si bien se tiene una muestra pequeña de datos (N=30), el proceso de minería de datos es válido ya que ésta representa la totalidad de los residuos orgánicos en 30 restaurantes en Acapulco por las facilidades que dieron los restauranteros para la recolección de datos, y “se emplean herramientas y aplicaciones que realizan análisis estadísticos sobre los datos” [11]. En este sentido, para el proceso de minería de datos se empleará el software de RapidMiner [5], el cual provee una interfaz gráfica de usuario muy amigable que permite realizar cruces de información y generar modelos estadísticos de manera visual, ya sea por un analista o bien por los mismos tomadores de decisiones. Este software se puede descargar de manera gratuita desde su sitio de Internet, tiene una versión libre que puede ser utilizada para fines académicos y de investigación, además es multiplataforma (funciona en distintos sistemas operativos); lo único que requiere tener preinstalado es la máquina virtual de Java, y que al momento de iniciarse por primera vez se instalen todos sus componentes desde el Internet.

3 Metodología

La aplicación de la metodología se realizará en una base de datos creada durante el proceso de trabajo de campo, aplicado a los residuos de los restaurantes en Acapulco, la muestra fue seleccionada por el método de conveniencia, derivado de las facilidades que otorgaron los responsables de la administración de los restaurantes. Se aplicó la metodología CRISP con apego a las etapas que la conforman.

La importancia de la base de datos como estrategia de captura y análisis de información a través de la minería de datos. También la utilización de la minería de datos en las actividades cotidianas de las organizaciones en general y de las empresas en particular, se ha convertido en algo cotidiano y más frecuente.

El objetivo de esta investigación, es construir una base de datos a través de un proceso de recolección de datos relacionada con los residuos orgánicos de los restaurantes del puerto de Acapulco, para desarrollar una metodología que transforme los datos en conocimiento y ese conocimiento en acción, para la toma de decisiones, teniendo la expresión máxima:

$$\text{Datos} + \text{Análisis} = \text{Conocimiento}$$

En [12] la metodología que se desarrolla en este trabajo involucra tres componentes relacionados que permiten llegar a la solución de nuestro problema, siendo estos: 1. La Base de Datos construida, 2. La Metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) por ser una de las metodologías más utilizadas en minería de datos y 3. La Herramienta de RapidMiner para el análisis de información, la cual nos permitirá visualizar los resultados de forma más eficiente.

En [13], menciona que RapidMiner es una herramienta óptima para crear modelos y a posterior la realización de análisis predictivos de grandes volúmenes de datos, facilitando el análisis con una avanzada analítica empleando solamente las ventajas o características más usadas, y a veces poco conocidas, en todas las versiones en Microsoft Windows, es aquella que se llama “drag and drop” y que simplemente quiere decir “arrastrar y soltar” y opcionalmente la generación de código, utilizándola también en aplicaciones como “analizar el impacto de ambiental de los residuos orgánicos”. Con esta herramienta se podrá:

- Tener acceso a los datos
- Crear un conjunto de datos óptimo para el análisis predictivo
- Descubrir patrones o problemas de datos
- Modelado, entre otras funcionalidades
- Estadísticas, Matriz de correlación, Gráficas y otros reportes de nuestros datos

Así al plantear el proyecto lo primero que debemos establecer son las preguntas del objetivo de la investigación, las preguntas que se plantean son:

- ¿Cuál es la estrategia adecuada que permita la toma de decisiones cimentadas en la base de datos obtenida de los residuos orgánicos de los restaurantes?

- ¿Cómo beneficia el uso de las herramientas informáticas a la toma de decisiones enfocadas en los impactos ambientales?

Al no existir datos recientes que permitan conocer la situación actual del estudio del puerto de Acapulco relacionado con la generación de residuos orgánicos por las empresas restauranteras, se elabora una cedula de campo para obtener datos primarios. La utilización de RapidMiner nos permitió el análisis de los resultados de manera gráfica, facilita su entendimiento y en todo caso nos provee de información en tiempo real que facilita la toma de decisiones de acuerdo a [10] esta se lleva a cabo en seis etapas:

- 1.- Distinguir y definir el problema
- 2.- Analizar a profundidad el problema
- 3.- Generar la mayor cantidad de alternativas de solución.
- 4.- Elegir la alternativa que mejor se adapte a la solución del problema.
- 5.- Elaborar el plan de acción de dicha alternativa.
- 6.- Llevar a cabo el plan anterior.

En esta metodología utilizamos herramientas informáticas en la captura y el análisis de información estratégica, la interpretación que la misma pueda generar son enriquecedoras, carece de la posibilidad de un análisis a profundidad en tiempo real. Para solventar esta carencia proponemos el uso de la herramienta de minería de datos visual [6].

La metodología CRISP-DM, está dividida en 4 niveles de abstracción organizados de forma jerárquica (Fig. 1) en tareas que van desde el nivel más general, hasta los casos más específicos y organiza el desarrollo del proyecto, en una serie de seis fases (Fig. 2):

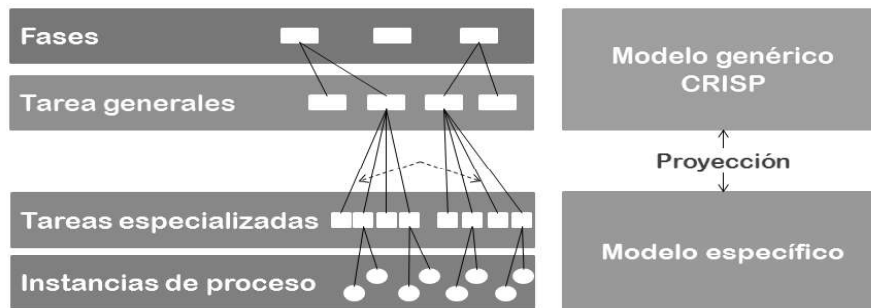


Fig. 1.
Esquema de los cuatro niveles de CRISP-DM.
Fuente: ([CRISP, 2000])

La sucesión de fases no es necesariamente rígida.

Cada fase es estructurada en varias tareas generales de segundo nivel. Las tareas generales se proyectan a tareas específicas, donde finalmente se describen las acciones que deben ser desarrolladas para situaciones específicas, pero en ningún momento se propone como realizarlas.

A continuación se describen cada una de las Fases en que se divide la metodología CRISP.

3.1 Fase de comprensión del problema

La falta de datos referentes a la cantidad actual de residuos que se generan en el municipio de Acapulco es alarmante, así como al poco interés que muestran las autoridades por el tema, debido a la poca información y a la consideración de otros sectores como más importantes, dejando de lado el de la gestión integral de los residuos.

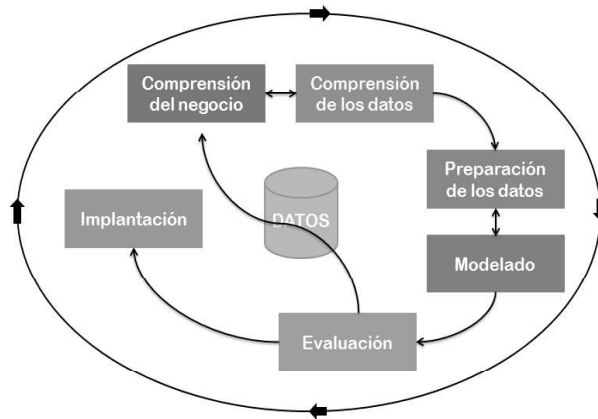


Fig. 2. Modelo del Proceso CRISP-DM Fuente: ([CRISP, 2000]).

Esta metodología se utiliza para obtener información referente a la generación de residuos orgánicos de los restaurantes de Acapulco, puesto que siendo Acapulco un lugar turístico que opera con una demanda estacionaria se reciben millones de turistas al año y no se tiene información sobre la generación de residuos orgánicos. Derivado de que no existe hasta el momento una base de datos que permita conocer en forma más específica la cantidad de residuos orgánicos que se generan por las empresas del ramo restaurantero se desarrolló una base de datos que permite tener una información actualizada que muestra las ventajas por utilizar las herramientas informáticas las cuales permiten una toma de decisiones asertiva en el aprovechamiento de la generación de residuos orgánicos así como ubicar los variables impactos ambientales que pueden derivar de éstos [14].

3.2 Fase de comprensión de los datos

Los datos que se recolectaron fueron correspondientes a 30 restaurantes con los cuales se obtuvo un análisis a fondo de la situación actual en cuanto a la generación de residuos de los restaurantes de Acapulco. La recolección de datos se dividió en 2 secciones, la primera fue la aplicación de cuestionarios, la cual se realizó del día 09 de noviembre al 26 de noviembre de 2016, contenían un total de 13 preguntas, que abarcaban desde la ubicación, hasta el consumo mensual promedio de luz. La segunda parte fue un pesaje de los residuos generados en el transcurso del día, esta segunda fase se realizó del 28 de noviembre al 03 de diciembre de 2016

3.3 Fase de preparación de datos

En esta fase se hicieron algunas adaptaciones necesarias para que RapidMiner pudiera procesar la base de datos y así obtener los resultados de manera gráfica.

Para que la herramienta de RapidMiner analizara los resultados, se formaron dos bases con los mismos datos de forma diferentes, por restaurantes y por variables, siendo éstas un total de 16.

3.4 Fase de modelado

Continuando con el proceso de la metodología de CRISP-DM e implementándola en RapidMiner describimos el paso de la fase de modelado:

Paso 1. Dentro de la pantalla principal, colocarse en **New Process** posteriormente seleccionar **Blank** y

Paso 2. En seguida la pestaña Add Data y seleccionar la carpeta donde se guardaran las bases de datos, en este caso My Computer.

Paso 3. Ahora se seleccionan las bases de datos y se presiona **Next**.

Paso 4. RapidMiner mostrará los datos contenidos en la base, seleccionando **Next**, posteriormente hará un análisis de los mismos.

Paso 5. Finalmente las bases quedan cargadas para poder trabajar en ellas, con una representación gráfica de pastel.

3.5 Fase de evaluación

En esta etapa se evalúa el modelo obtenido que da respuesta a objetivo planteado en este proyecto de minería de datos y el funcionamiento de RapidMiner sobre la base de datos de la investigación.

3.6 Fase de implementación

Una vez ingresada la base de datos en RapidMiner se realizó un análisis de los resultados, obteniendo un modelo resultado del proceso del algoritmo escogido en la herramienta informática y haciendo énfasis en los datos para que no existiera alguna modificación de los mismos.

3.6.1 Determinación de los próximos pasos

Una vez revisadas las bases de datos se determinó que los resultados que arrojaban eran factibles para la investigación, además de que RapidMiner presenta la posibilidad de seguir alimentando la base para futuros resultados. Las gráficas que presenta RapidMiner son las más factibles para la investigación

4 Resultados experimentales

La utilización de herramientas como RapidMiner permitió analizar de forma gráfica los resultados obtenidos durante el trabajo de campo. La interfaz gráfica de esta herramienta mejoró el entendimiento de los datos obtenidos durante el trabajo de campo, lo que agilizó la toma de decisiones.



Fig. 3. Análisis de resultados con RapidMiner.

De acuerdo a los resultados obtenidos en esta investigación se analiza que de 30 restaurantes distribuidos en las 3 zonas más importantes de Acapulco (costera Miguel Alemán, Caleta y Caletilla y finalmente Pie de la Cuesta), las cuales representan las zonas con más afluencia de turistas durante las diferentes temporadas del año. El 56.6% de los restaurantes están categorizados como de 2 tenedores, es decir, son el tipo de restaurantes con precios accesibles a todo tipo de bolsillos. Se observa además que el 98% de los restaurantes son micro empresas, que cuentan con un total máximo de 10 empleados, los cuales atienden en promedio en temporada baja hasta 30 comensales por día y en temporada alta hasta 120 comensales por día.

En cuanto a la separación de basura, el 66.6% de los restaurantes no hacen una separación de residuos, sin embargo el 76.6% del total de la muestra consideran que la reutilización de los residuos orgánicos son muy importantes, sobre todo si estos son transformados en algún subproducto del mismo que permita un bien común, la opción ideal sería la transformación a energía eléctrica, que representa uno de los costos más importante para ellos, pagando un promedio mensual de \$5,000 pesos en temporada baja y hasta \$8,000 pesos en temporada alta. Este costo aumenta por las adaptaciones que deben hacer al restaurante en temporada alta tales como instalación de aires acondicionados, ventiladores, refrigeradores extras, entre otros.

El promedio de generación en temporada baja de residuos es de 13.89 kg, de los cuales 7.64 kg son orgánico y 6.25kg son inorgánicos, entonces si Acapulco cuenta con un total de 4,078 restaurantes, en total por el sector restaurantero se estarían generando en promedio 56,643.42 kg de residuos al día, logrando llegar al cuádruple en temporada alta (226,573.68kg). De estos totales, los residuos orgánicos promedios generados al día en temporada baja equivalen a 31,155.92 kg., mientras que en temporada alta son 124,623.68 kg, sin una base de datos que se alimente y pueda arrojar información a los tomadores de decisión al momento de establecer una estrategia de aprovechamiento de los residuos orgánicos, estos terminarían en el relleno sanitario de Acapulco sin ser tratados para un posible beneficio a la industria.

5 Conclusiones y Trabajo Futuro

El uso de la minería de datos en el análisis de bases de datos permite analizar a profundidad los resultados y ser utilizados por softwares especializados. La utilización de programas especializados en base de datos amplía el conocimiento de las personas inexpertas del tema, en ciertas ocasiones solamente se utiliza Excel para capturar información sin embargo en este caso el uso de este software facilitó la creación de la base de datos.

La interfaz gráfica de RapidMiner permite la comprensión y el análisis a profundidad de los resultados arrojados por la base de datos a través de la estadística, gráficas, matriz de correlación entre los datos y algunas otras características de la base de datos. Al mostrar resultados gráficamente, permite que la toma de decisiones sea más rápida y más precisa y al alimentar constantemente esta base de datos, permitirá que la toma de decisiones sea constante, adaptable y modificable, de acuerdo a las necesidades que se presenten. Derivado de la utilización de la base de datos se puede establecer una relación entre la cantidad de residuos generados en Acapulco y el potencial aprovechamiento de estos de tal manera que se establezcan medidas de concientización por parte de la localidad de la importancia de tener información en tiempo real que les permita establecer medidas de control y aprovechamiento de los residuos que les permita mantenerse como un destino turístico atractivo a los turistas promoviendo la imagen de sustentabilidad en el destino.

Agradecimientos. Queremos agradecer a los revisores por sus acertados comentarios, a la COFAA-IPN por su apoyo en el desarrollo de esta investigación y a las autoridades de la oficina de turismo del municipio de Acapulco, Gro por las facilidades para la recolección de datos.

Referencias

- [1] Jaramillo, G., & Zapata, L. (2008). *Aprovechamiento de los residuos sólidos orgánicos en Colombia*. Universidad de Antioquia. Recuperado a partir de <http://uniciencia.ambientalex.info/infoCT/Apressolorgco.pdf>
- [2] Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, Rüdiger Wirth. CRISP-DM 1.0, Step-by-step Data Mining Guide, CRISP-DM, 2000
- [3] Oldemar Rodríguez Rojas. *Metodología para el Desarrollo de Proyectos en Minería de Datos* CRISP-DM, 2010.
- [4] Durand, M. (2011). *La gestión de los residuos sólidos en los países en desarrollo: ¿cómo obtener beneficios de las dificultades actuales?* Espacio y Desarrollo, (23), 115–130.
- [5] RapidMiner. (2015). *Herramienta Informática*. Florida International University, USA. consultado el 29 de Agosto de 2015, disponible en: <http://rapid-i.com/content/view/181/190>.
- [6] RapidMiner. <http://rapid-i.com/content/view/181/196/lang,en/> [Rapid-I, 2011]
- [7] Román, J. (2015). *Inteligencia en Redes de Comunicaciones, Minería de Datos*: consultado el 20 de Octubre de 2016, disponible en: <http://ocw.uc3m.es/ingenieria-telematica/inteligencia>.
- [8] Monserrat, S. (2015). *Minería de Datos en Base de Datos de Servicios de Salud*. Universidad Tecnológica Nacional - Facultad Regional Santa Fe, Centro de Investigación y Desarrollo de Ingeniería en Sistemas de Información" (CIDISI).
- [9] Lei Li, Chao Shen, Long Wang, et al. (2014). *iMiner: Mining Inventory Data for Intelligent Management*. USA. Florida International University, School of Computing and Information Sciences.
- [10] Rodríguez, M., Márquez, M. (2015). *Manejo de problemas y toma de decisiones* (2a ed., Vol. 8). México, DF: El Manual Moderno S.A. de C.V. Recuperado a partir de https://mgptercera2014.files.wordpress.com/2014/01/manejo_de_problemas_y_toma_de_decisiones_vol_8_2a_ed_22_to_39.pdf
- [11] Pérez M. (2014). *Minería de datos / Data mining: La Metodología Crisp-dm De IBM*. El Lenguaje CLEM E IBM Spss Modeler / the IBM Crisp-dm Methodology. the CLEM Language and IBM Spss Modeler.
- [12] Pérez, M. (2015). *Minería de Datos a través de ejemplos*. España: RC Libros.
- [13] Salazar, R. (2014). *IBM Business Analytics: Minería De Datos*. IBM Spss Modeler. USA. Createspace Independent Publishing Platform.
- [14] Perdomo, C. (2008). *Técnicas para el aprovechamiento y valorización de los residuos orgánicos con alta tasa de biodegradación*. Recuperado el 27 de agosto de 2016, a partir de <ftp://ftp.ani.gov.co/Aeropuertos/Sur%20Occidente/T%C3%A9cnico/5.%20INVERSIONES>.