

# Modelos econométricos y minería de datos: cómo se pueden relacionar\*

Juan Sebastián Upegui Villegas\*\*

## RESUMEN

La minería de datos se ha posicionado en las últimas décadas como un enfoque que presenta grandes posibilidades de aplicación en dominios de datos diversos. Uno de los campos de estudio donde la minería de datos aporta valor y es de gran importancia es en la gestión de conocimiento a nivel empresarial. En este orden de ideas, en este trabajo se quiere presentar una reflexión sobre cómo se pueden relacionar los modelos econométricos con la minería de datos y su aplicación para dar respuesta a preguntas asociadas, por ejemplo, a cómo financiar pequeñas y medianas empresas (pymes) en Colombia. El análisis de variables como el crecimiento de la pyme, las tasas de interés de un banco comercial, tasas de interés de un banco de segundo piso y la tasa de reinversión o rentabilidad financiera (ROE) de la empresa puede ser abordado por medio de la construcción de modelos estadísticos o econométricos, sin embargo, es de interés revisar cómo la minería de datos podría contribuir a dicho análisis.

*Palabras clave:* minería de datos; modelo econométrico; pyme; tasa de interés; tasa de reinversión.

---

\* Semillero específico en Minería de Datos, Facultad de Ingenierías, 2021-1

\*\* Estudiante de Ingeniería Financiera, Universidad de Medellín, Semillero en Minería de datos, Correo electrónico: villegas.uegui@gmail.com

## *Econometric Models and Data Mining: How they Can Be Related*

### **ABSTRACT**

Data mining has positioned itself in the last decades as an approach with great potential for diverse data domains. One of the fields of study where data mining is of great importance is the management of knowledge at the enterprise level. In this order of ideas, this article aims towards presenting a reflection on how econometric models can be related to data mining and their application to answer questions related, for example, to how to finance small and medium-sized enterprises (SMEs) in Colombia. The analysis of variables such as SME growth, commercial bank interest rates, interest rates, and the reinvestment rate or return on equity (ROE) of the enterprise can be approached through the construction of statistical or econometric models, however, it results interesting to first review how data mining could contribute to such analysis.

*Keywords:* data mining, econometric model, smes, rate of returns, reinvestment rate.

## INTRODUCCIÓN

Las pequeñas y medianas empresas, pymes, son un importante renglón en la generación de empleo en el país (Vera-Colina et al., 2014), por lo cual el análisis de las posibilidades de financiación que mejor se adecuen a sus necesidades y condiciones particulares es un aspecto de interés. El análisis de estas posibilidades puede ser realizado, luego de seleccionar una serie de variables, por medio de modelos estadísticos o econométricos que permitan entender las correlaciones entre dichas variables y soportar la selección de soluciones ante problemas relacionados con la financiación de las pymes. Los directivos de las pymes afrontan retos constantes en términos de revisar las mejores opciones a la hora de conseguir capital para dar soporte a la operación de sus proyectos y con ello poder aportar a la mejora de la economía con mayor inversión, generación de empleos y oportunidades para todos.

La teoría de jerarquía financiera postula que los costos de financiación aumentan cuando la información es asimétrica (Myers, 1984). Los recursos provienen de tres fuentes principalmente: fondos internos, deuda y nuevo capital. Las empresas priorizan sus fuentes de financiación, primero prefiriendo financiación interna, y luego, deuda, y finalmente incrementando capital como última opción para conseguir recursos. De acuerdo a esto, la financiación interna es usada primero; cuando esta se agota, entonces la deuda es emitida y cuando ya no es razonable emitir más deuda, se emite capital. Esta teoría mantiene que los negocios se adhieren a una jerarquía de fuentes de financiamiento y prefieren la financiación interna cuando se dispone, y la deuda es preferida sobre nuevo capital si la financiación externa es requerida. Entonces, la forma de deuda que una empresa escoge puede actuar como señal de su necesidad para financiamiento externo (Myers y Majluf, 1984).

Al analizar problemas de financiación, la selección de variables explicativas es uno de los aspectos cruciales, y se considera como uno de los problemas importantes en estadística y econometría. En la construcción de modelos econométricos se puede caer en problemas de sobre parametrización, lo que da como resultado un buen desempeño a nivel intra-muestral, pero no en la misma medida para los casos extra-muestrales. Así, la selección de las variables explicativas que representen de forma adecuada a la variable dependiente es una labor que requiere especial cuidado y hace parte de las tareas de preparación que deben sufrir los datos, previo a la aplicación de alguna técnica de análisis.

Reconociendo que la minería de datos se ha venido consolidando como una herramienta de análisis potente y que se puede acoplar a los procesos de gestión del conocimiento a nivel empresarial, se considera que algunas de las técnicas que hacen parte de este enfoque se podrían usar para relacionar factores clave, detectar categorías, pronosticar escenarios, entre otras. Por ello, en este artículo se desea

presentar una reflexión sobre cómo podrían relacionarse las técnicas de minería de datos con los modelos econométricos asociados a problemáticas como el análisis de fuentes de financiación para el crecimiento de las pymes. En el desarrollo de este documento se conceptualizan los enfoques involucrados en la problemática que se está analizando, se tratará de responder a algunas preguntas clave y, finalmente, se concluirá con las consideraciones finales.

## 1. MODELOS ECONOMÉTRICOS

Según Maddala (1996) "La econometría es la aplicación de métodos estadísticos y matemáticos al análisis de datos económicos con el propósito de dar contenido empírico a las teorías económicas y verificarlas o refutarlas" (p. 25).

Los modelos econométricos son aplicados a las variables económicas y financieras para realizar predicciones en las mismas variables económicas actuales, y son muy útiles para;

- El análisis estructural: ayuda a entender el funcionamiento de la economía.
- La predicción de los valores: pronosticar el futuro a través de variables económicas.
- La simulación para planificar las posibilidades que pueden ofrecer las variables establecidas en el modelo.
- La simulación para controlar los valores óptimos que logramos hallar en las variables utilizadas para aplicar en el entorno analizado en el modelo.

## 2. MINERÍA DE DATOS

Según López (2007), la minería de datos puede definirse inicialmente como "un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos" (p. 3). La minería de datos es una técnica de análisis de datos que surge para dar respuesta a la disponibilidad, cada vez mayor, de datos almacenados bajo conjuntos estructurados en bases de datos o datos semiestructurados o sin estructura, pero que crecen cada día más.

Como se ha indicado, la extracción de patrones, tendencias y regularidades para entender y describir mejor los datos es la función u objetivo de la minería de datos. Sin embargo, para llegar a ella, se vivió un proceso de evolución de los sistemas de información y de la forma de utilización de los datos. Es decir, antes de llegar a la minería de datos se fueron adoptando algunos enfoques y tecnologías que permitieron llegar a los análisis actuales, esto de la mano del crecimiento constante en la cantidad de datos producidos y almacenados.

La minería de datos ostenta amplia aplicabilidad en campos de estudio; en general se podría decir que desde que se cuente con datos, la minería puede ser aplicada; sin embargo, no se debe olvidar que cada objetivo exige diferentes técnicas. A continuación, según Oliveira y Da Silva (2009) y Riquelme, Ruiz, y Gilbert (2006) se presentan algunas áreas en las cuales la minería de datos ha sido aplicada de forma satisfactoria:

- Retención de clientes: se identifican los perfiles para determinados productos.
- Bancos: identificación de patrones para apoyar la gestión de relaciones con los clientes.
- Tarjetas de crédito: identificación de segmentos de mercado, patrones de rotación.
- Cobranza: detección de fraudes.
- *Telemarketing*: facilidad de acceso a información del cliente.
- Electoral: identificación de perfiles de posibles votantes.
- Medicina: refinamiento de diagnósticos más precisos.
- Seguridad: detección de actividades terroristas y criminales.
- Investigaciones biométricas.
- Toma de decisiones: filtro de información relevante para fortalecer bases como indicadores de probabilidad.
- Educación: estudio de los perfiles estudiantiles y comportamientos en los procesos académicos.
- Agencias de viaje: aumento en el volumen de ventas dirigiendo paquetes a clientes según perfiles.
- Farmacia: efectividad de los tratamientos.
- Astronomía: identificación de nuevas estrellas y galaxias.
- Agricultura: identificación de áreas para uso en diferentes cultivos.
- Ciencias ambientales: identificación de modelos de funcionamiento de ecosistemas naturales.
- Ciencias sociales: estudio de flujos de la opinión pública, planificación urbana.

Cabe resaltar que no son estas todas las áreas de aplicabilidad, sino un ejemplo de algunas entre muchas otras. Esta amplia aplicabilidad puede verse explicada en que la minería de datos combina o acoge un conjunto de técnicas y herramientas que dan soporte al análisis de diferentes tipos de datos y que se pueden también ajustar y acoplar con otras técnicas de análisis o enfoques para la construcción de modelos o extracción de patrones (Pérez Marqués, 2014). No obstante, es importante que la selección de las técnicas de minería de datos a usar se ajuste a las realidades de los datos y a los objetivos perseguidos e identificados en cada problemática particular.

### **3. IMPORTANCIA DEL ANÁLISIS DE LAS PYMES EN EL PAÍS**

Colombia cuenta con un potencial económico, el cual, en gran medida, es gracias a las pymes. En la actualidad, Colombia tiene 2.540.953 de mipymes, que representan el 90 % de las empresas del país, producen el 30 % del PIB y emplean más del 65 % de la fuerza laboral nacional (Redacción Economía, 2020). Pero el 62 % de estas empresas no cuenta con acceso a fuentes de capital externas; en otras palabras, 1.575.391 empresas no tienen opciones para financiar sus proyectos los cuales son vitales para generar empleo. De allí la necesidad de indicar a las pymes del país cuáles son otras opciones para conseguir este capital, ya que la falta de información y la dificultad en el acceso a la misma hacen que las empresas no lleguen a conocer los programas especiales que les ayudarían a conseguir dicho capital.

### **4. RELACIÓN ENTRE LA MINERÍA DE DATOS Y LOS MODELOS ECONÓMICOS PARA EL ANÁLISIS DE LA FINANCIACIÓN DE PYMES**

Algunos trabajos previos encontrados en la literatura dan una orientación sobre la relación que puede tener la minería de datos en temas relacionados con la financiación y los modelos econométricos. En Marulanda Echeverry *et al.* (2017), se realizó un estudio en el cual se utilizó la minería de datos como enfoque para la gestión de conocimiento de pymes en Colombia, y se encontró que las pymes requieren avanzar en la implementación de buenas prácticas para la gestión del conocimiento relacionadas con la generación, retención y compartir el conocimiento.

Por su parte, Fernández-Rodríguez *et al.* (2009) realizaron un análisis y propusieron una metodología para la aplicación de minería de datos en la selección de modelos econométricos para casos en los que se cuenta con varios candidatos para las variables explicativas o regresores y con una elevada correlación entre sí; en concreto, se utilizaron algoritmos genéticos para hacer dicha selección, validando la propuesta en la selección de variables que explican el crecimiento económico, la predicción del fracaso empresarial y la selección de activos financieros en una cartera que sigue un índice bursátil. Los autores logran concluir que existe una amplia capacidad en el uso de la minería de datos para la selección de modelos con potencial de aplicación en numerosos campos de la economía.

En cuanto a las oportunidades que ofrece contar con mecanismos para hacer análisis de los datos generados en las pymes, Cárdenas Olaya (2020) señala que entre los beneficios que pueden obtener las pymes al contar con herramientas de análisis de datos se encuentran la mejora de procesos, la reducción de costos, la obtención de resultados esperados y el apoyo a la toma de decisiones. Se cuenta también con trabajos en los que se ha analizado la aplicación de minería de datos para análisis de datos relacionados con estrategias de *marketing* para pymes (Escobar Terán *et al.*, 2016), análisis de factores de fracaso (Arquero Montaña *et al.*, 2009) e, incluso, identificación de patrones que influyen en la calidad de planes de negocios a apoyar por programas de emprendimiento, así como apoyo de microempresarios (Reyes Pérez *et al.*, 2017).

En relación con lo anterior, una de las posibilidades que se vislumbra en cuanto a la relación o complemento que se puede dar entre la minería de datos y los modelos econométricos es en la selección y clasificación de los datos, dado que la minería de datos puede permitir mejoras en estas dos tareas y con ello también optimizar el desempeño del modelo econométrico, en términos de precisión y confiabilidad. Además, las técnicas de minería permiten, por medio de la aplicación de algoritmos, conocer y comprender o explicar la distribución de los datos; de esta manera, se podría, incluso, llegar a detectar datos con valores atípicos e identificar las variables o atributos que estén quitando valor o agregando valor dentro de los modelos de regresión. Al relacionar estas dos disciplinas podemos conseguir datos más confiables y mirar la variabilidad de los mismos, lo que da como resultado que el modelo que se construya sea mucho más ajustable a las condiciones de la vida real.

## **CONCLUSIONES**

Colombia se encuentra en un proceso de adopción de políticas de datos abiertos y reglamentación en temas asociados al uso, distribución y manejo de datos. Actualmente el acceso a bases de datos del sector financiero y económico es limitado, lo cual restringe las posibilidades para llevar a cabo análisis y modelos asociados a variables que deben ser extraídas de este dominio de datos. Esto repercute a la hora de tratar de generar un modelo econométrico donde las variables explicativas sean la tasa de interés del banco comercial, la tasa de interés de un banco estatal y la tasa de reinversión de las pymes, con la tasa de crecimiento de las pymes como variable dependiente. Se considera que estas variables deben ser seleccionadas, dado que son las principales formas de conseguir financiación para una pyme.

Asimismo, se debe considerar que la información que se puede extraer de los mercados no es perfecta, y en el caso de Colombia, dicha información es aún más asimétrica, acompañado esto de la mencionada dificultad para tener acceso a datos que son manejados solo por las grandes empresas, los bancos y el Gobierno.

Si se contara con mayor conocimiento sobre las opciones para financiar las pymes del país se podría contribuir a tener una tasa de supervivencia también mayor. Si la mayoría de las personas que desean emprender o están emprendiendo, conocen la realidad del contexto financiero en el que se tendrán que ubicar, aumentaría su capacidad de acoplamiento y planeación de los costos de financiación.

La minería de datos es un enfoque con aplicabilidad promisoría en temas relacionados con la generación de modelos econométricos para el análisis de problemáticas asociadas a la financiación de pymes en Colombia, así como para el análisis de factores de crecimiento de las mismas.

## BIBLIOGRAFÍA

- Arquero Montañó, J. L., Abad Navarro, M. C., & Jiménez Cardoso, S. M. (2009). Procesos de fracaso empresarial en pymes. Identificación y contrastación empírica. *Revista Internacional de La Pequeña y Mediana Empresa*, 1(2), 64–77. <https://idus.us.es/handle/11441/28503>
- Cárdenas Olaya, D. S. (2020). Importancia de un análisis de datos en las pymes. <https://repository.unimilitar.edu.co/handle/10654/37743>
- Escobar Terán, H. E., Alcívar, M., y Puris, A. (2016). Aplicaciones de minería de datos en *marketing*. *Revista Publicando*, 3(8), 503–512. <https://dialnet.unirioja.es/servlet/articulo?codigo=5833425&info=resumen&idioma=SPA>
- Fernández-Rodríguez, F., Acosta-González, E., y Andrada-Félix, J. (2009). Especificación de modelos econométricos utilizando minería de datos. *Revista Electrónica de Comunicaciones y Trabajos de ASEPUMA*, 10, 223–252. <https://dialnet.unirioja.es/servlet/articulo?codigo=3267309&info=resumen&idioma=ENG>
- López, C. P. (2007). Minería de datos: técnicas y herramientas. In *Paraninfo*. *Paraninfo*. [https://books.google.com.pe/books?id=wz-D\\_8uPFCEC&printsec=frontcover&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.com.pe/books?id=wz-D_8uPFCEC&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false)
- Maddala, G. (1996). *Introducción a la econometría*. Prentice Hall.
- Marulanda Echeverry, C. E., López Trujillo, M., y Mejía Salazar, M. H. (2017). Minería de datos en gestión del conocimiento de pymes de Colombia. *Revista Virtual Universidad Católica del Norte*, 50, 224–237. <https://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/821>
- Myers, S. C. (1984). The Capital Structure Puzzle. *The Journal of Finance*, 39(3), 574–592. <https://doi.org/10.1111/j.1540-6261.1984.tb03646.x>
- Myers, S. C., y Majluf, N. S. (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics*, 13(2), 187–221. [https://doi.org/10.1016/0304-405X\(84\)90023-0](https://doi.org/10.1016/0304-405X(84)90023-0)
- Oliveira, C., y Da Silva, J. C. (2009). *Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas*. [www.inf.ufg.br](http://www.inf.ufg.br)
- Pérez Marqués, M. (2014). *Minería de datos a través de ejemplos* (1aed.). RC Libros.
- Redacción Economía*. (2020, September 8). El 62 % de las pymes colombianas no tiene acceso a financiamiento. Colombia Fintech. <https://www.colombiafintech.co/novedades/el-62-de-las-pymes-colombianas-no-tiene-acceso-a-financiamiento>

- Reyes Pérez, I. F., Castañeda Caycedo, M., Ariza Zambrano, A., y Rodríguez Martínez, R. M. (2017). *Aplicación de minería de datos para identificar patrones que influyen en la calidad de la formulación de planes de negocio de emprendimiento* Sena - Regional Distrito Capital. <https://alejandria.poligran.edu.co/handle/10823/1134>
- Riquelme, J. C., Ruiz, R., y Gilbert, K. (2006). Minería de datos: conceptos y tendencias. *Revista Iberoamericana de Inteligencia Artificial*, 29, 11–18. <http://www.aepia.org>
- Vera-Colina, M. A., Melgarejo-Molina, Z. A., y Mora-Riapira, E. H. (2014). Access to finance for Colombian SMES: A look at financial indicators. *Innovar*, 24(53), 149–160. <https://doi.org/10.15446/innovar.v24n53.43922>