

**RInCE-Revista de Investigaciones del Departamento de Ciencias
Económicas de La Universidad Nacional de La Matanza**

Recensión bibliográfica:

Data mining for the masses

Autores: Ángel Héctor de Mendonça, Alba Ester Iribarne y Alan Yurquina
Becario de investigación. Departamento de Ciencias Económicas. UNLaM

Título de la obra reseñada: Data mining for the masses

Título del capítulo (si fuese una compilación): Chapter one: Introduction
to Data Mining and CRISP-DM; Chapter Four: Correlation.

Apellido y nombre de los autores de la obra: Matthew North.

Editorial: Global Text

Nº de edición: Segunda edición

Año de edición: 2012

Lugar de edición: Estados Unidos de América (USA)

Número de ISBN: 0615684378

Formato: electrónico

URL de acceso:

<https://docs.rapidminer.com/downloads/DataMiningForTheMasses.pdf>

Fecha de acceso: 12/10/2017

Desarrollo de la recensión de la obra

Chapter One: Introduction to Data Mining and CRISP-DM

En su obra, Matthew North pretende describir las principales características y conceptos de la disciplina data mining y de cómo a través de la vida cotidiana se generan datos que hacen posible un análisis de los fenómenos que permitirían tomar decisiones más informadas en orden del cumplimiento de objetivos. Nos parece pertinente incorporar este texto a nuestra investigación ya que está dirigido tanto a estudiantes de grado como profesionales que no poseen un profundo conocimiento de data mining. En la obra se utiliza un software con una interfaz de usuario amigable llamado *Rapidminer*, aunque sí es necesario tener conocimiento sobre conceptos y métodos estadísticos básicos.

De esta forma, el autor propone acertadamente un diagrama metodológico al momento de realizar un análisis de la información. Como primera instancia propone un entendimiento organizacional, es decir entender que es lo se construirá en base a las preguntas que necesitan respuesta. Define un segundo paso como entendimiento de los datos, dónde se determinarán las fuentes de obtención de los mismos y se procurará que la información sea confiable y representativa. Una tercera etapa consiste en la homogeneización de los datos que procurará un resultado confiable. La cuarta y quinta etapa constan en la construcción del modelo computarizado que se conforma de la aplicación de algoritmos que identifican y muestran los patrones existentes en la información; de este modo, estos modelos permitirán realizar clasificaciones y predicciones dando lugar a la última etapa. Finalmente, el modelo conceptual obtenido formará parte de la información disponible para tomar decisiones o entender el funcionamiento de un sistema.

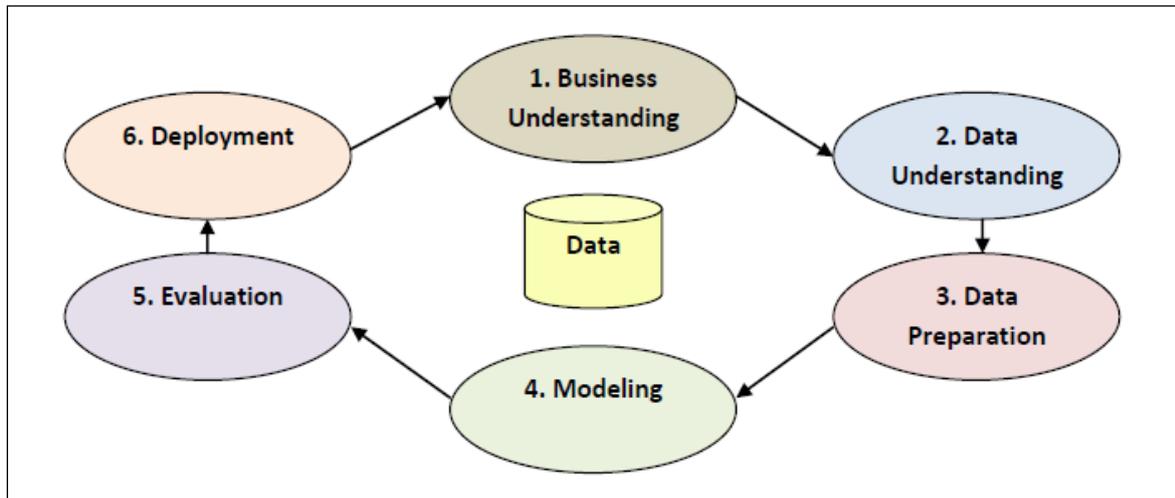


Gráfico 1: Modelo Conceptual. Elaboración del autor Matthew North

Chapter Four: Correlation

Una correlación es una medida de que tan fuerte es la relación entre dos o más variables donde el coeficiente varía entre 1 y -1. La distinción del signo tiene que ver con la dirección del movimiento entre las dos variables correlacionadas; cuando el coeficiente es un valor positivo, significa que cuando una de variables se eleva, la otra también se eleva en cambio cuando el coeficiente es negativo, cuando una variable aumenta la otra disminuye.

Data mining for the masses

Ángel Héctor de Mendonça, Alba Ester Iribarne y Alan Yurquina

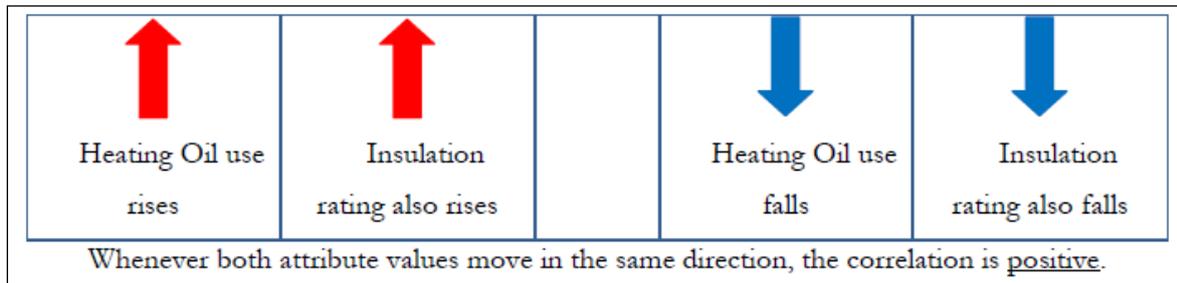


Gráfico 2: Ejemplo correlaciones positivas. Elaboración del autor.

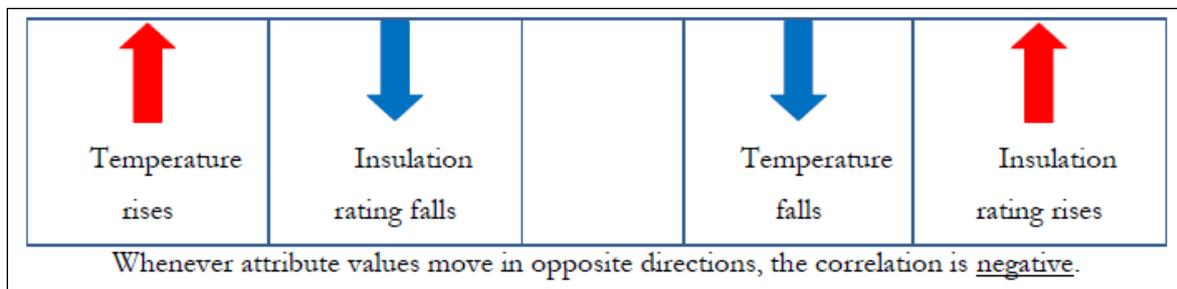


Gráfico 3: Ejemplo correlaciones negativas. Elaboración del autor.

Del mismo modo, mientras más cerca de 1 se encuentre el coeficiente, más fuerte es la relación entre ellas.

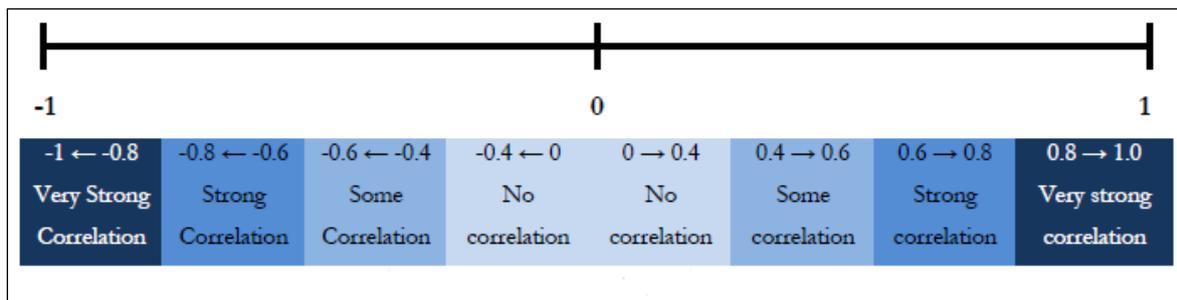


Gráfico 4: Escala intensidad de correlaciones. Elaborado por el autor.

Los coeficientes calculados a través del modelo expresan una relación, pero no una causalidad entre variables. Por lo tanto, deben reconocerse y aceptarse las limitaciones del modelo de correlaciones de manera de no llegar a conclusiones e interpretaciones erróneas de los hechos.

Consideraciones finales

Encontramos relevante para nuestra investigación “*Influencia Sobre La Composición De La Estructura Económica Argentina De Las Políticas Del Estado Nacional En El Periodo 2003-2011*”¹ dirigida por Ángel Héctor De Mendonça y Alba Ester Iribarne en calidad de Co-Directora, la búsqueda de correlaciones cómo resulta expuesto en el capítulo 4 de este libro, en el cual el autor logra ejemplificar exitosamente la aplicación de este método de modelización, a través de situaciones sencillas. Destacamos la relevancia de su utilización como herramienta de aprendizaje para los estudiantes de la carrera de Licenciatura en Economía.

¹“Influencia Sobre La Composición De La Estructura Económica Argentina De Las Políticas Del Estado Nacional En El Periodo 2003-2011”, N° C2-ECO-043. Director del proyecto: Ángel Héctor de Mendonça, Co-Director del proyecto: Alba Ester Iribarne, Integrantes del equipo: Prof. Claudia Andrada, Prof. Clara Cinquegrani, Prof. Sonia Re, Alumno Lucas Pittito (Becario UNLAM), Alumno Alan Yurquina (Becario UNLAM).