

TÉCNICAS DE PREDICCIÓN MEDIANTE MINERÍA DE DATOS EN LA INDUSTRIA ALIMENTARIA BAJO EL PARADIGMA DE INDUSTRIA 4.0

Guerrero, Manuel; Luque, Amalia*; Lama, Juan Ramón

Grupo TEP022

Diseño Industrial e Ingeniería del Proyecto y la Innovación. Escuela Politécnica Superior. Universidad de Sevilla.

*E-mail: aluque2@us.es

RESUMEN

Un objetivo del paradigma de Industria 4.0 es convertir los datos en información útil para la toma de decisiones. Tras el término de minería de datos (Data Mining) se engloba un conjunto de técnicas dirigidas a la extracción de conocimiento procesable y que se encuentra implícito en los datos analizados. Las técnicas de minería de datos se emplean para mejorar el rendimiento de procesos productivos o industriales en los que se manejan grandes volúmenes de información estructurada. Las técnicas de predicción se basan en el análisis de datos para predecir funciones de valores continuos. La industria alimentaria es una industria de especial relevancia en nuestro país. Poder ajustar el proceso de producción al volumen de demanda, tipo de materia prima, composición y a las condiciones de funcionamiento de la fábrica permitirá tomar decisiones que lleven a la reducción de costes, a la mejora del control y a la reducción de las incidencias técnicas con clientes. En este trabajo se mostrarán como las principales técnicas de predicción pueden ser aplicadas a la industria alimentaria, en aras de avanzar hacia los objetivos que propone la Industria 4.0.

PALABRAS CLAVE

Industria alimentaria, Industria 4.0, Técnicas de predicción, Minería de datos.

ABSTRACT

One objective of the Industry 4.0 paradigm is to convert data into useful information for decision making. After the term of Data Mining, a set of techniques is included, aimed at the extraction of processable knowledge and that is implicit in the analyzed data. Data Mining techniques are used to improve the performance of productive or industrial processes in which large volumes of structured information are handled. Prediction techniques are based on data analysis to predict continuous value functions. The food industry is an industry of special relevance in our country. Being able to adjust the production process to the volume of demand, type of raw material, composition and the operating conditions of the factory will allow decisions to be taken that reduce costs, improve control and reduce technical incidents with customers. In this paper we will show how the main prediction techniques can be applied to the food industry, to advance towards the objectives proposed by Industry 4.0.

KEYWORDS

Food industry, Industry 4.0., Prediction techniques, Data mining.

INTRODUCCIÓN Y OBJETIVOS

En cualquier aplicación de Minería de Datos nos encontramos con el problema de diseño de la aplicación, procesamiento y transformación de datos que dan como resultado una representación del problema. El objetivo último es convertir los datos en información útil para la toma de decisiones.

El procesamiento de datos apunta a abordar una serie de problemas, como la redundancia de datos, la inconsistencia, el ruido, la heterogeneidad, la transformación, el etiquetado, el desequilibrio de datos y la representación/selección de características [1]. La preparación y el procesamiento de los datos suelen ser costosas, debido al requerimiento, presencia e intervención de un humano y una gran cantidad de opciones para elegir.

Este trabajo aborda los desafíos que se deben afrontar al hacer uso de la Minería de datos que más tarde será de utilizada para diversas técnicas predictivas.

METODOLOGÍA

Para la elaboración de este trabajo se han utilizado fuentes internacionales como Scopus, Web of Science y ScienceDirect; para buscar literatura relacionada. Con la literatura relevante encontrada, se ha realizado un análisis para determinar el papel de la Minería de Datos en la Industria 4.0.

En los siguientes apartados se pueden encontrar procesos por los que deben pasar los datos para poder ser utilizado y diferentes técnicas predictivas. Por último, una aplicación práctica para industria de alimentaria.

RESULTADOS Y DISCUSIÓN

La preparación y el procesamiento de los datos suelen ser costosas, debido al requerimiento, presencia e intervención de un humano y una gran cantidad de opciones para elegir. Además, algunos supuestos de datos convencionales no son válidos para el Data Mining, por lo que algunos métodos tradicionales de procesamiento se vuelven inviables.

Por otro lado, la minería de datos crea la oportunidad de reducir la dependencia de la supervisión humana al obtener información de fuentes de datos extensas, diversas y de transmisión directa.

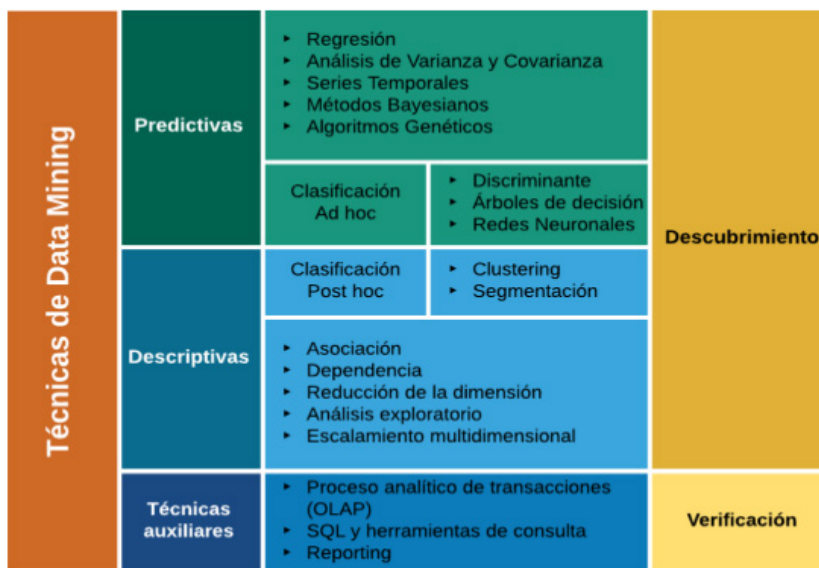


Figura 1. Técnicas de Minería de Datos.

Los datos deben ser transformados en información, deben ser tratados antes de poder hacerse uso de ellos. La Minería de datos nos proporciona las herramientas para gestionarlos.

Redundancia de datos

La duplicación surge cuando dos o más muestras de datos representan la misma entidad. El impacto de la duplicación de datos o la inconsistencia puede ser grave [2].

Ruido de datos

La falta de valores, valores incorrectos, dispersión de datos y valores atípicos pueden introducir ruido en el tratamiento de los mismos [3].

Heterogeneidad de datos

Las bases de datos actuales ofrecen datos de múltiples vistas y diferentes tipos de repositorios, en formatos dispares y de diferentes muestras de la población y, por lo tanto, muy heterogéneos [4].

Discretización de datos

Algunos algoritmos, como los árboles de decisión, solo pueden tratar con atributos discretos. La discretización convierte los datos cuantitativos en datos cualitativos, obteniendo una división no superpuesta de un dominio continuo [5].

Datos desequilibrados

El problema de los datos desequilibrados se ha abordado mediante métodos tradicionales de muestreo aleatorio y estratificado. Sin embargo, el proceso puede llevar mucho tiempo si involucra iteraciones en la generación de submuestras y cálculo de métricas de error [6].

Técnicas Predictivas

Con las técnicas predictivas se pretenden obtener información valiosa para la toma de decisiones. Existen multitud de herramientas, aquí presentamos algunas de ellas:

Red neuronal artificial (ANN, Artificial Neural Network)

Una red neuronal es una red de neuronas no lineales interconectadas inspiradas en los estudios del sistema de neuronas de organismos biológicos. Una red neuronal produce un patrón de salida cuando se presenta con un patrón de entrada. Tiene una arquitectura distribuida en paralelo con un gran número de nodos (neuronas) y conexiones [7].

Algoritmo genético (GA, Genetic Algorithm)

Es una forma de técnicas de investigación de función común adaptativas motivadas por la evolución natural sugerida por Holland [8]. Pretenden modelar el procedimiento de evolución en la naturaleza. Cada cromosoma consta de varios genes y presenta el resultado candidato a la dificultad presentada. El cruce produce dos nuevos cromosomas al cruzar dos elementos de la población. La mutación cambia arbitrariamente los genes de los individuos [9].

Árbol de regresión

Es una herramienta de análisis que puede ser utilizada para clasificar, predecir, interpretar y manipular datos [10]. Características: simplifica relaciones complejas, relativamente fácil de interpretar, se construye con un enfoque no requiere el cumplimiento de supuestos estadísticos, maneja distribuciones sesgadas sin la necesidad de realizar transformaciones sobre los datos y es robusta a las puntuaciones extremas [11].

Los árboles ayudan a buscar subgrupos específicos y relaciones que tal vez no se encontrarían con los estadísticos tradicionales; de esta manera, ayudan a revelar información oculta [12].

Método ARIMA

La metodología ARIMA (Autorregresivos Integrados de Medias Móviles) fue desarrollada en 1970 [16]. Tienen la ventaja que es usualmente plausible encontrar un proceso que proporcione una descripción adecuada de los datos [13].

La metodología involucra un proceso iterativo [14]: análisis de Estacionalidad, identificación del Modelo, estimación de Parámetros del Modelo, verificación del Modelo, pronóstico y validación del pronóstico.

Modelos de Alisado Exponencial

Son procedimientos para ajustar una curva apropiada a datos históricos de una determinada serie de tiempo. Hasta hace relativamente poco tiempo aun no se desarrollaba un procedimiento que incorporara una estructura de modelado [15].

Regresión multivariable

Más conocida como Multivariate Adaptive Regression Splines (MARS), pretende aproximar de manera eficiente la relación de más de dos variables para la obtención de una variable de predicción [16]. Se utiliza cuando se estudia la posible relación entre varias variables independientes entre sí (variables predictoras o explicativas) y otra variable dependiente (variable de respuesta).

Sector Agroalimentario: alimentación y bebidas.

El objetivo de este trabajo es señalar la relevancia que pueden tener la Minería de Datos y técnicas predictivas como herramientas que favorezcan la gestión de la producción en el sector agroalimentario, más concretamente en el sector de alimentación y bebidas.

Fruto del crecimiento de las industrias de alimentos y bebidas nos encontramos un sector especialmente exigente con la gestión de la producción: comportamiento económico (oferta rígida, demanda inelástica), intervención administrativa en las producciones primarias, requerimientos en seguridad alimentaria, diversidad de productos, variabilidad de las producciones... a las que hay que añadir las que también afectan al resto de la economía: fluctuaciones financieras, aceleración de las tecnologías aplicables, globalización de los mercados... [17]

La industria alimentación y bebidas es un sector estratégico muy importante para nuestra economía. Genera más de medio millón de empleos directos, por encima del total de la industria manufacturera y del conjunto de la economía española [18]. Por

este motivo se ha centrado el caso de estudio en este sector, debido al gran potencial que muestran estas empresas y las posibilidades que facilitan.



Figura 2. Tratamiento de datos para determinar patrones.

CONCLUSIONES

Gracias al nuevo paradigma de Industria 4.0, haciendo uso de la tecnología de la información, ahora es más fácil convertir los datos en información útil para la toma de decisiones.

Herramientas como la minería de datos y técnicas predictivas las organizaciones disponen de información que les ayuda a plantear estrategias empresariales más precisas, efectivas y aplicables en periodos de tiempos más cortos.

Antes de poder interpretar los datos de los procesos, las empresas deben tratarlos para reducir el problema a tratar y optimizar los recursos disponibles. La Minería de datos proporciona los medios para el tratamiento de los datos productivos.

Por otro lado, existen distintas técnicas predictivas, cada una con sus bondades y deficiencias, que proporcionan una nueva vía a las empresas para la toma de decisiones.

La industria agroalimentaria española es susceptible de beneficiarse del paradigma de Industria 4.0. En un mercado globalizado, cualquier ventaja pequeña respecto a la competencia puede marcar una gran diferencia.

Siendo el mercado de la alimentación y bebidas muy exigente respecto a la gestión de la producción, las herramientas que se han discutido más arriba ofrecen a las organizaciones una ventaja competitiva a valorar.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Bengio, Y., Courville, A., y Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), pp. 1798-1828.
- [2] Chen, Q., Zobel, J., y Verspoor, K. (2015). Evaluation of a machine learning duplicate detection method for bioinformatics databases. En *Proceedings of the ACM Ninth International Workshop on Data and Text Mining in Biomedical Informatics* (pp. 4-12). ACM.
- [3] Cao, L., Wei, M., Yang, D., y Rundensteiner, E. A. (2015). Online outlier exploration over large datasets. En *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 89-98). ACM.
- [4] Gandomi, A., y Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), pp. 137-144.
- [5] Ramírez S., García, S., Mouriñon, H., Martínez, D., Bolón V., Alonso, A., y Herrera, F. (2016). Data discretization: taxonomy and big data challenge. Wiley Interdisciplinary Reviews: *Data Mining and Knowledge Discovery*, 6(1), pp. 5-21.
- [6] Su, Y., Agrawal, G., Woodring, J., Myers, K., Wendelberger, J., y Ahrens, J. (2014). Effective and efficient data sampling using bitmap indices. *Cluster computing*, 17(4), pp. 1081-1100.
- [7] Lakshmanan, I., y Ramasamy, S. (2015). An artificial neural-network approach to software reliability growth modeling. *Procedia Computer Science*, 57, pp. 695-702.
- [8] Holland John, H. (1975). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. USA: University of Michigan.
- [9] Kalavathi, J., Balamurali, S., y Venkatesulu, M. (2015). An efficient evolutionary approach for identifying evolving groups in dynamic social networks using genetic modeling. *Procedia Computer Science*, 57, p. 428.
- [10] Song, Y. Y., y Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives*, 27, p. 130.

- [11] Hernández, R. V. R., Quintero, J. M. M., Mendoza, A. L., y Almazán, D. Á. (2018). Identificación de perfiles en la satisfacción de los usuarios de repositorios digitales a través de un árbol de regresión. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 9(17), pp. 1-19.
- [12] Berlanga, S. V., Rubio, H. M. y Vilá Baños, R. (2013). Cómo aplicar árboles de decisión en SPSS. *REIRE Revista d'Innovació i Recerca in Educació*, 6(1), pp. 65-79.
- [13] Box, G., y Jenkins, G. (1970). *Time Series Analysis-Forecasting and Control*. San Francisco: Holden Day. p. 553.
- [14] Hyndman, R., Koehler, A. B., Ord, J. K., y Snyder, R. D. (2008). Forecasting with exponential smoothing: the state space approach. *Springer Science & Business Media*.
- [15] Ord, K., y Lowe, S. (1996). Automatic forecasting. *The American Statistician*, 50(1), pp. 88-94.
- [16] Ramírez, A. M. A. (2013). *Métodos utilizados para el pronóstico de demanda de energía eléctrica en sistemas de distribución*. Universidad Tecnológica de Pereira.
- [17] Ministerio de Agricultura, Pesca y Alimentación. (2018). Marco Estratégico para la Industria de Alimentación y Bebidas. Recuperado de: <https://www.mapa.gob.es/es/alimentacion/temas/industria-agroalimentaria/marco-estrategico/>
- [18] Federación Española de Industrias de la Alimentación y Bebidas (FIAB). (2018). Sector alimentación y bebidas. Recuperado de: <http://fiab.es/sector/>