



XXIII Edición del Workshop de Investigadores en Ciencias de la Computación

LIBRO de ACTAS

Abril 2021 - Chilecito – La Rioja – Argentina

Universidad Nacional de Chilecito - UNdeC

Red de Universidades con Carreras de Informática - RedUNCI

CIENCIA DE DATOS APLICADA. ESTUDIO DE CASOS EN DIVERSAS ÁREAS, FOCALIZADO EN LA INDUSTRIA AGROPECUARIA DE LA PROVINCIA DE RIO NEGRO

Paola Britos¹, Paola Pizzingrilli^{1,6}, Lina María Montoya Suárez³, Pamela Britos⁴, Gastón Di Bonis², Melisa Isaja², Giuliana Fois¹, Gustavo Agüero¹, Pablo Enrique Argañarás^{1,5}, Martín René Vilugrón^{1,5}, Maximiliano Donadio¹

¹ Universidad Nacional de Río Negro. Laboratorio de Informática Aplicada. Río Negro. Argentina. pbritos@unrn.edu.ar, ppizzingrilli@unrn.edu.ar, gfois@unrn.edu.ar, gustavo54ar@gmail.com, parganaras@unrn.edu.ar, maxdonadio21@gmail.com.

² Universidad Nacional de Río Negro. Río Negro. Argentina. melisa.isaja@gmail.com, gdibonis@unrn.edu.ar

³ Universidad Católica Luis Amigó, Grupo de Investigación SISCO
Medellín, Colombia lina.montoyasu@amigo.edu.co

⁴ Sede Regional Sur: Metán – Rosario de la Frontera, Universidad Nacional de Salta. Argentina. pamebritos@gmail.com

⁵ Comisión Nacional de Energía Atómica, Centro Atómico Bariloche, banda@cab.cnea.gov.ar, martinvilu@cab.cnea.gov.ar

⁶ Universidad Nacional de Río Negro. Instituto de Investigaciones de Recursos Naturales, Agroecología y Desarrollo Rural (IRNAD). Río Negro. Argentina. ppizzingrilli@unrn.edu.ar

RESUMEN

Las organizaciones están buscando una nueva manera de producir optimizando y potenciando sus recursos, es por ello que están incorporando tecnologías que conectan el mundo físico con el virtual a través de diversos dispositivos que le permitan acceder a grandes cantidades de datos en tiempo real y tomar mejores decisiones en todo el proceso productivo. Por otra parte, la Ciencia de Datos es un conjunto de principios fundamentales que apoyan y guían la extracción de información y conocimiento a partir de los datos; incluye diversas metodologías, técnicas, algoritmos y herramientas que facilitan el procesamiento avanzado y automático de los mismos; permitiendo identificar información relevante y estratégica, que a simple vista no es detectada. Este proyecto tiene por *objetivo articular integradamente mecanismos de proceso de interpretación de grandes masas de información en diversas industrias, (con hincapié en la agropecuaria de la provincia de Río Negro), a través diversas técnicas de tratamiento de datos y su visualización brindada por la Ciencia de Datos.*

Palabras claves: Data Science aplicada, explotación de información, aplicaciones agropecuarias.

CONTEXTO

Este proyecto de I&D se encuentra radicado en la Universidad Nacional de Río Negro. Cuenta con los siguientes antecedentes:

- Proyectos de investigación:
 - Estudio y Evaluación de Tecnologías de la Información y la Comunicación para el desarrollo de ciudades inteligentes en Río Negro (PI 40-C-270).
 - Explotación y visualización de información. Su aplicación a las industrias rionegrinas (PI 40-C-542).

1. RELEVANCIA DEL PROBLEMA

Las organizaciones están buscando una nueva manera de producir optimizando y potenciando sus recursos, es por ello que están incorporando tecnologías que conectan el mundo físico con el virtual a través de diversos dispositivos que le permitan acceder a grandes cantidades de datos en tiempo real y tomar mejores decisiones en todo el proceso productivo: desde la concepción del producto o servicio hasta y comercialización del mismo, el estudio de los datos de las organizaciones es realizado por la Ciencia de Datos.

Por otra parte, de acuerdo al Boletín Oficial de la República Argentina [1] “el sector científico y tecnológico tiene el potencial de contribuir gracias a sus aportes e innovaciones, a elevar la competitividad y productividad de la economía, del país, impactando en el mejoramiento de la

calidad de vida de la población...”. Cabe destacar, que una ciudad latinoamericana que se destaca en este tipo de innovación, es Medellín, se mantiene a la vanguardia de este tipo de tecnología, ya que ha sabido planificar y adoptar tecnologías en diversos ambientes de la ciudad por las que el mundo la ha premiado de manera regular como una ciudad innovadora, sustentable e inteligente [2] y que a través de diversas actividades en la región [3], [4] y [5] y en conjunto con diversas organizaciones, como por ejemplo Instituto Nacional de Tecnología Agropecuaria (INTA Río Negro) se han realizado innovaciones al respecto.

Por lo anteriormente citado, es que el objetivo es *articular integradamente mecanismos de proceso de interpretación de grandes masas de información de diversos tipos de industria, haciendo hincapié en la agronómica, a través diversas técnicas de tratamiento de datos y su visualización, las cuales son brindada por la Ciencia de Datos.*

1. INTRODUCCIÓN

1.1. Cambio en las Organizaciones

De acuerdo a lo informado por [1] uno de los ejes estratégicos del Gobierno Nacional Argentino es impulsar procesos de transformación productiva que estimulen el desarrollo económico, promoviendo las capacidades tecnológicas y productivas de las empresas, la creación de empleos de calidad, la participación en los mercados internacionales, la mejora de la productividad y la incorporación y el desarrollo de las tecnologías emergentes.

El surgimiento de nuevos desafíos en el ámbito de las organizaciones está siendo afectado transversalmente por transformaciones tecnológicas y productivas que tienen impacto en la dinámica económica de los países que resultan determinantes para su desarrollo, planteando nuevos retos y demandas productivas [6], [7]. Las tecnologías que se insertan hoy en las organizaciones tienen que ver con la automatización, la inteligencia artificial, el *uso de ciencia de datos*, entre otras [4], [5], [3].

Ciencia de Datos

Un estudio realizado en [8] menciona que cada vez son más las empresas que se basan en datos, las cuales han mejorado su desempeño en medidas objetivas de los resultados financieros y operativos. En un entorno tan competitivo como el actual, las organizaciones, de cualquier tipo, requieren estrategias para poder adaptarse a los cambios y generar conocimiento a partir de los datos disponibles, es una muy buena estrategia [9]. Debido a la evolución de las TICs, los sistemas de almacenamiento pueden acopiar gran cantidad de datos, de diferente índole, a un menudo coste; lo que conlleva a que estos crezcan de manera exponencial en pequeños intervalos de tiempo. Los datos no son recogidos únicamente para el mantenimiento de registros o para validar las hipótesis generadas por los recursos humanos, sino que a medida que el volumen de datos crece, existe la posibilidad de probar las hipótesis que aún no se habían previsto [10]. El crecimiento en cantidad y diversidad de los datos ha llevado a que las herramientas convencionales de administración práctica, no puedan manejar estos conjuntos de datos [11].

Para poder tomar decisiones pertinentes y administrar estos conjuntos de datos potencialmente invaluable, es necesario realizar un análisis exhaustivo sobre los mismos. La Ciencia de Datos o Data Science (DS) es un conjunto de principios fundamentales que apoyan y guían la extracción de información y conocimiento a partir de los datos; incluye diversas metodologías, técnicas, algoritmos y herramientas que facilitan el procesamiento avanzado y automático de los mismos; permitiendo identificar información relevante y estratégica, que a simple vista no es detectada [9], [11], [12]. Fois, Aguero y Britos [13] desde la perspectiva de la ciencia de datos realizan una comparación de metodologías ágiles para aplicar en proyectos de ciencias de datos, del que se han tomado las actividades más inherentes para planificar cómo abordar el problema y organizar las actividades para abordar este caso de estudio. En síntesis, puede definirse la DS, como la aplicación de métodos cualitativos y cuantitativos para predecir resultados y solucionar problemas relevantes las cuales debido a la creciente e inmensa cantidad de datos que existe actualmente, el conocimiento y el análisis del dominio no pueden estar separadas. Existen muchos

algoritmos para el procesamiento y extracción de datos, así como una gran cantidad de detalles sobre los métodos aplicados en este campo; pero DS implica mucho más que solo algoritmos de Data Mining. DS proporciona a los profesionales del área una estructura y un conjunto principios que brindan un marco para tratar sistemáticamente los problemas de extracción de conocimiento útil de los datos; donde los métodos y la metodología para tratar los datos y llevar adelante este tipo de proyectos son trascendentales [11], [12]. Dentro de la ingeniería de sistemas, los proyectos de DS son un tipo especial de proyectos; cuyo objetivo es extraer conocimiento relevante, idóneo y original de los datos [14], [15].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Este proyecto tiene por objetivo principal articular integradamente mecanismos de proceso de interpretación de grandes masas de información, de diversos tipos de industria (con hincapié en la industria agropecuaria y la educación de la provincia de Río Negro – Argentina, y Medellín - Colombia), a través diversas técnicas de tratamiento de datos y su visualización brindada por la Ciencia de Datos.

Específicamente tiene por objetivo: (a) Identificar modelos de comportamiento de crecimiento en diversos entornos, teniendo presente las características del mismo, y (b) Implementar nuevos modelos de visualización y estudio de datos brindados con la Industria y los estudios científicos de diversos centros académicos, en diferentes temáticas.

3. RESULTADOS OBTENIDOS/ ESPERADOS

A continuación, se indican los resultados obtenidos hasta el momento:

- Melisa Isaja, Paola Pizzingrilli, Pamela Britos, Maximiliano Donadio, Giuliana Fois, Gustavo Agüero, Pablo Enrique Argañarás, Martín René Vilugrón, Lina María Montoya Suárez, Paola Britos y Gastón Di Bonis. 2020. Estudio de comportamiento de cultivo de mostaza blanca (*Sinapis alba L.*). Un caso de aplicación con Big Data. XI Congreso

Internacional sobre Aplicación de Tecnologías de la Información y Comunicaciones Avanzadas (ATICA 2020). ISBN 978-84-18254-84-0.

- Viviana Moscher, Paola Britos. Desarrollo de un modelo predictivo para descubrir factores que inciden en la deserción de alumno en la Facultad de Humanidades, Universidad Nacional del Nordeste. XI Congreso Internacional sobre Aplicación de Tecnologías de la Información y Comunicaciones Avanzadas (ATICA 2020). ISBN 978-84-18254-84-0
- Caleb Faillance, Florencia Pollo, Britos. Análisis De Sentimientos: Detección de Sarcasmo en Español a Través de Emoticones En Twitter. XI Congreso Internacional sobre Aplicación de Tecnologías de la Información y Comunicaciones Avanzadas (ATICA 2020). ISBN 978-84-18254-84-0.
- Victor Rojo, Florencia Pollo, Britos. Desarrollo de un recurso léxico de palabras informales en español de Argentina para el análisis de sentimientos en Twitter. XI Congreso Internacional sobre Aplicación de Tecnologías de la Información y Comunicaciones Avanzadas (ATICA 2020). ISBN 978-84-18254-84-0.
- Melisa Isaja, Paola Pizzingrilli, Pamela Britos, Maximiliano Donadio, Giuliana Fois, Gustavo Agüero, Pablo Enrique Argañarás, Martín René Vilugrón, Lina María Montoya Suárez, Paola Britos y Gastón Di Bonis. 2020. BIG DATA, aplicado a la Agricultura. IX Semana de la Tecnología, Creatividad e Innovación. Universidad Luis Amigo. Medellín. Colombia.
- Britos, P; Fois, Guiliana; Agüero Cobella, G. 2020. Evaluación comparativa de las metodologías Team Data Science Process TDSP y Analytics Solutions Unified Method for Data Mining ASUM-DM desde la perspectiva de la ciencia de datos. Investigación Formativa en Ingeniería. 264-270. ISBN 978-958-52-3335. Medellín. Colombia.
- Britos, Paola; Montoya Suarez, Lina María; Vilugron, M.; Donadio, Maximiliano. Ciencia de Datos en

Ingeniería de Sistemas Virtual y Distancia de la Uniremington. XXIII Encuentro Nacional de Investigación 2020. Universidad Luis Amigo. Medellín. Colombia.

4. FORMACION DE RECURSOS HUMANOS

Este proyecto tiene por objetivo la formación de: 3 tesis de maestría, 2 tesis de grado, 4 investigadores juniors.

5. BIBLIOGRAFÍA

- [1] M. de P. y Trabajo, “Boletín Oficial de la República Argentina,” *RESFC-2019-1-APN-MPYT*. Feb. 22, 2021, [Online]. Available: <https://www.boletinoficial.gob.ar/#!DetalleNorma/208007/20190522>.
- [2] “Medellín, el milagro de una Ciudad Inteligente,” *PanamericanWorld*. Feb. 22, 2014, [Online]. Available: <https://panamericanworld.com/revista/economia/medellin-el-milagro-de-una-ciudad-inteligente/>.
- [3] P. Britos, “Proyecto de I&D: ‘Explotación y visualización de información. Su aplicación a las industrias rionegrinas.’” 2017.
- [4] P. Britos and H. Vivas, “Proyecto de I&D: ‘Estudio y Evaluación de Tecnologías de la Información y la Comunicación para el desarrollo de ciudades inteligentes en Río Negro.’” Universidad Nacional de Río Negro, 2015.
- [5] H. Vivas, “Proyecto de I&D: ‘Herramientas Informáticas para el Desarrollo de Servicios Digitales Innovadores para las Comunidades urbanas y Rurales en el Marco de Ciudades y Regiones Inteligentes.’” Universidad Nacional de Río Negro, 2017.
- [6] L. F. O. Clavijo, J. D. F. Ledesma, S. C. Nieto, and C. J. G. Duque, “Computación en la Nube, estudio de herramientas orientadas a la Industria 4.0,” *Lámpsakos*, no. 20, pp. 68–75, Feb. 2018, [Online]. Available: <https://dialnet.unirioja.es/servlet/articulo?codigo=6771316>.
- [7] N. Velásquez, E. C. Estévez, and P. M. Pesado, “Cloud Computing, Big Data and the Industry 4.0 Reference Architectures,” in *VI Jornadas de Cloud Computing & Big Data - JCC & BD 2018 (La Plata, junio 2018)*, Feb. 2018, vol. 18, n, [Online]. Available: <http://sedici.unlp.edu.ar/handle/10915/71687>.
- [8] E. McAfee, A., & Brynjolfsson, “Big Data: The Management Revolution | Request PDF,” *ResearchGate*. Feb. 22, 2012, [Online]. Available: https://www.researchgate.net/publication/232279314_Big_Data_The_Management_Revolution.
- [9] K. B. Eckert and P. V. Britos, “Modelo basado en la Toma Decisiones con Criterios Múltiples para la elección de metodologías de Data Science,” *XX Workshop de Investigación en Ciencias de la Computación*, p. 5, 2018.
- [10] R. Agarwal and V. Dhar, “Editorial-Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research,” *Inf. Syst. Res.*, vol. 25, no. 3, pp. 443–448, Feb. 2014, doi: 10.1287/isre.2014.0546.
- [11] M. A. Waller and S. E. Fawcett, “Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management,” *J. Bus. Logist.*, vol. 34, no. 2, pp. 77–84, Feb. 2013, doi: <https://doi.org/10.1111/jbl.12010>.
- [12] F. Provost and T. Fawcett, “Data Science and its Relationship to Big Data and Data-Driven Decision Making,” *Big Data*, vol. 1, no. 1, pp. 51–59, Feb. 2013, doi: 10.1089/big.2013.1508.
- [13] P. Fois, G.; Agüero, G.; Britos, “Evaluación comparativa de las metodologías Team Data Science Process TDSP y Analytics Solutions Unified Method for Data Mining ASUM-DM desde la perspectiva de la ciencia de datos,” *Investigación Formativa en Ingeniería*, vol. 4ta, pp. 264–270, 2020, doi: <https://zenodo.org/record/4031253#.YDTxX-gzbiU>.
- [14] P. Pytel, P. Britos, and R. García-Martínez, “A Proposal of Effort Estimation Method for Information Mining Projects Oriented to SMEs,” in *Enterprise Information Systems of the Future*, vol. 139, G. Poels, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 58–74.
- [15] J. Schiefer, J. Jeng, S. Kapoor, and P. Chowdhary, *Process Information Factory: A Data Management Approach for Enhancing Business Process Intelligence*.