

## Estado de madurez tecnológica de las PyMEs argentinas a partir de Ciencia de Datos.

Análisis comparativo de informes del INTI y datos abiertos nacionales.

Seminario, Andrea G.<sup>1,2</sup>; Lagier, María Eugenia<sup>2</sup>; Tornillo, Julián E.<sup>1</sup>; Pascal Guadalupe<sup>1</sup>

<sup>1</sup> Facultad de Ingeniería, UNLZ - <sup>2</sup> Instituto Nacional de Tecnología Industrial (INTI)

### RESUMEN

Actualmente la industria nacional busca herramientas para abordar la transformación digital, por lo que distintas organizaciones científico-tecnológicas generan diagnósticos para conocer cuál es el grado de madurez organizacional que presentan las PyMEs para entender sus necesidades.

El Instituto Nacional de Tecnología Industrial (INTI) evaluó el nivel de madurez de más de 600 Pymes Argentinas durante el año pasado. Estos documentos deben ser firmados de forma manual por el asesor evaluador ya que según esta caracterización se asignan los financiamientos que otorga la Secretaría de Industria y Desarrollo Productivo para los Programas de Apoyo a la Competitividad (PAC).

Este trabajo busca conocer las características de las PyMEs que participaron del PAC, cuáles son sus principales necesidades y qué tecnologías, tanto digitales como de gestión (*Kaizen*), pueden aportar las entidades anteriormente mencionadas. Para ello, se desarrolló un algoritmo en *Python* que extrae expresiones regulares de forma automática y masiva de los diagnósticos en PDF, los convierte en CSV y los analiza.

NIVEL	ACCIÓN
NIVEL 1	PREPARAR
NIVEL 2	IMPLEMENTAR
NIVEL 3	DESPLGAR
NIVEL 4	MEJORAR

### OBJETIVOS

- Identificar relaciones entre el nivel de madurez organizacional y la capacidad de adopción de tecnologías 4.0.
- Estudiar las distintas tecnologías 4.0 que pueden aplicarse en sectores industriales estratégicos.
- Extrapolar los resultados obtenidos al marco nacional usando las bases de datos abiertos del Registro MiPyME (+2M).
- Optimizar la metodología de análisis de datos a través de un algoritmo de código abierto.
- Aplicar técnicas de ciencia de datos y *web scraping* en documentos PDF.

### METODOLOGÍA

Marco de trabajo: **CRISP-DM** (*Cross Industry Standard Process for Data Mining*)

- Comprensión del problema y definición de objetivos (*outputs*).
- Comprensión del origen de los datos (*inputs*).
- Preparación de los datos.
- Modelado del algoritmo.**
- Retroalimentación.
- Se establece la metodología de implementación.

Elección de la técnica: **Web Scraping** adaptado a archivos PDF

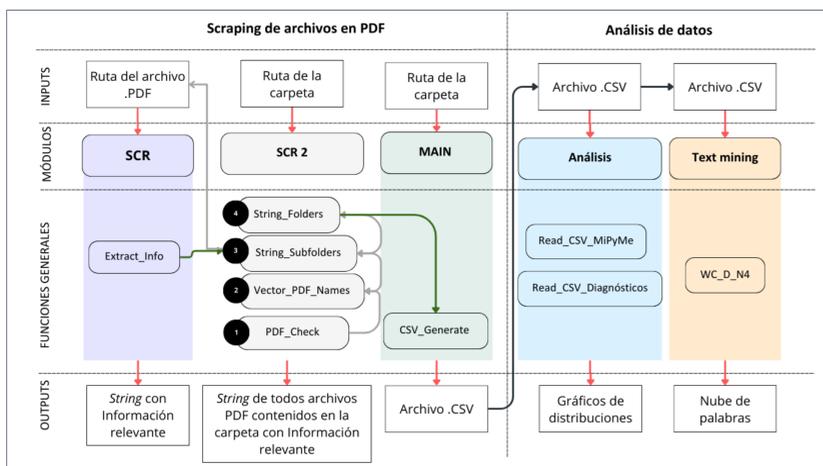
- Elección del Leguaje (Python por ser de código abierto).
- Elección de la función OCR (librería *PdfPlumber*).
- Definición de las expresiones regulares (Librería *Re*).
- Definir formato de salida (CSV).

Análisis de datos: Paquete *Pandas*.

### RESULTADOS

#### ALGORITMO

El algoritmo procesa el 95% de los informes de forma correcta en 80 segundos y los analiza en 13 segundos.



#### ANÁLISIS DE LA INDUSTRIA A TRAVÉS DE LOS DATOS DEL PAC

- Aproximadamente el 48% tienen niveles bajos de estandarización de procesos.
- Los sectores metalmeccánicos y automotrices cuentan mayores niveles de mejora continua.
- Más del 80% registran datos del proceso, pero el 43,68% no los analiza.
- Más del 40% toma los datos mensualmente mientras que el 11,35% los toma en tiempo real.



#### CONTRIBUCIÓN AL ESTADO DEL ARTE

- "Niveles de *Kaizen* vs Transformación Digital", publicación del departamento de Tecnologías de Gestión – INTI.
- Estado de madurez tecnológica de las PyMEs argentinas a partir de Ciencia de Datos, publicado en el Simposio de Informática en el Estado de la JAIIO 52.

### PRÓXIMOS DESAFÍOS

- Identificar capacidades para la adopción de tecnologías 4.0 a nivel nacional.
- Cuantificar la calidad de los datos extraídos.
- Modelar las expresiones regulares de los informes finales y analizarlas, usando NLTK (*Natural Language Tool Kit* – Paquete de *Python* que utiliza IA).
- Usar algoritmos de IA para el procesamiento de imágenes en archivos PDF.

### CIENCIA ABIERTA



52 JAIIO



#### AGRADECIMIENTOS

- Becas cofinanciadas INTI-UNLZ.
- Proyecto Lomas CyT.
- Equipo de Mejora de Procesos (14 FI UNLZ).