Estado de madurez tecnológica de las PyMEs argentinas a partir de ciencia de datos.

Análisis comparativo de informes del INTI y datos abiertos nacionales.

Andrea G. Seminario[1,2], Julián E. Tornillo[1], María Eugenia Lagier[2] y Guadalupe Pascal[1]

[1] UNLZ – Facultad de Ingeniería, Ruta 4 Km. 2, Buenos Aires, Argentina
{jtornillo, gpascal} @ingenieria.unlz.edu.ar
[2] INTI – Depto. de Tecnologías de Gestión, Av. Gral. Paz 5445, Buenos Aires, Argentina
{aseminario , mlagier}@inti.gob.ar

Resumen. El Departamento de Tecnologías de Gestión del Instituto Nacional de Tecnología Industrial brinda acompañamiento a las PyMEs argentinas con el objetivo de impulsar su competitividad adoptando herramientas de la filosofia japonesa *Kaizen*. Para ello, se documenta el estado de la organización al inicio y al final para evaluar los resultados obtenidos.
Este trabajo busca generar nuevos conocimientos a partir de estos documentos que permitan comprender el panorama general de las empresas para la generación herramientas e instrumentos que respondan a sus necesidades.
Se procesan y analizan los documentos generados para los programas de Apoyo a la Competitividad 2020-2021 y Productividad 4.0 2021-2022 y se contrastan con más de dos millones de datos abiertos del Ministerio de Economía utilizando técnicas y herramientas de ciencia de datos. Para ello se desarrollan dos algoritmos en Python utilizando Pandas, Matplotlib, Pdf Plumber y Re: el primero usa la técnica *text scraping* para la extracción automática y masiva de la información de los documentos en PDF y el segundo permite la homologación y comparación de estos con los datos abiertos.
Los resultados muestran los niveles de madurez de las PyMEs argentinas participantes y permite identificar capacidades para la adopción de tecnologías 4.0.

Keywords: Ciencia de datos, Datos abiertos, PyMEs, Industria 4.0, Mejora Continua.

1 Introducción

El actual paradigma 4.0 brinda nuevos sets tecnológicos al conjunto de la cadena de valor desde los menos hasta los más informatizados debido al flujo creciente de datos tratados que se convierten en información útil para la toma de decisiones y la elección de estrategias competitivas.[1]

En Argentina, el Plan de Desarrollo Productivo Argentina 4.0 (2021) es el programa del Estado Nacional que brinda soporte a las organizaciones para la incorporación del paradigma 4.0 desde el financiamiento en infraestructura y tecnología hasta la capacitación y asesoría técnica para el desarrollo de soluciones 4.0 brindadas por el Instituto Nacional de Tecnología Industrial (INTI). [2]

Este trabajo se desprende de la investigación que lleva a cabo el departamento de Tecnologías de Gestión del INTI para entender la madurez digital en las organizaciones. Para ello, la institución ha desarrollado un prediagnóstico basado en modelos teóricos de medición del desempeño, la mejora continua y la filosofía *Kaizen*[3], ya que se sabe, que su aplicación y fortalecimiento de esta cultura aumentan la productividad en las organizaciones. Con esta herramienta se miden los niveles de *Kaizen* en las organizaciones para entender su grado de madurez con relación a la mejora continua y lograr proporcionarles las herramientas adecuadas según su grado de conocimiento. [4]

La solución que brinda este trabajo se relaciona con las prácticas que se llevan actualmente en el almacenamiento de la información ya que se sabe que es una problemática recurrente en las instituciones[5] sobre todo si requieren de espacios de almacenamiento en formato de texto, es por ello, que se modelan 2 algoritmos en Python que posibilitan la extracción y análisis de los datos almacenados en los diagnósticos y los informes en formato PDF.

Vale aclarar que estos documentos deben estar firmados de forma manual o virtualmente por los Asesores de Tecnologías de Gestión (ATG) del INTI ya que certifican el estado registrado tanto al inicio como al final de la asistencia, por lo que el formato PDF ha sido el elegido por el departamento.

La importancia del análisis de estos datos radica en las tendencias de clase mundial hacia la digitalización [6] y en cómo el aporte de presupuesto Nacional puede favorecer a las organizaciones que elijan maximizar su competitividad a través de la mejora continua. En otras palabras, hace posible responder a las siguientes preguntas: ¿Cómo es el panorama general de las empresas que se postulan respecto a la mejora continua? ¿La mejora continua acentúa las oportunidades de adopción de tecnología 4.0? ¿Se puede considerar que los programas nacionales son aprovechados de forma federal? ¿Los esfuerzos estatales están contribuyendo al aumento de la competitividad de las organizaciones?

Actualmente estas preguntas no pueden ser respondidas de forma automática sin tener que analizar cada uno de los documentos elaborados por empresa para cada programa. Por lo que este trabajo además de analizar los datos extraídos busca ser una solución escalable y automática que pueda procesar grandes volúmenes de diagnósticos e informes finales de las PyMEs acompañadas por INTI.

Este procesamiento de datos permite el análisis de distribuciones sectoriales y provinciales a partir de los datos abiertos extraídos de Datos Argentina (Ministerio de Economía), [7] específicamente los registros MiPyME que cuentan con más de dos millones de datos recolectados de forma federal. Esta comparativa permitiría generar un primer paneo exploratorio sobre la representatividad de los resultados obtenidos en las empresas que participan en los programas a nivel nacional.

2 Desarrollo

2.1 Marco de trabajo

Este trabajo se construye sobre las bases de la ciencia abierta en consonancia con el movimiento de datos abiertos y los programas de apoyo a la competitividad del estado nacional[8], es decir, que busca generar conocimiento libre y gratuito con el objetivo de hacer más accesible y ágil la resolución de problemas relacionados a la extracción de datos de documentos PDF.

Actualmente se conocen diferentes marcos de trabajo en Ciencia de Datos con diferentes enforques y manteniendo una metodología de trabajo común. Particularmente en este trabajo, se utiliza la metodología *Cross Industry Standard Process for Data Mining* (CRISP-DM) que propone un ciclo de vida similar al ciclo PDCA utilizado para gestión de la mejora continua en los procesos de análisis de datos.[9] Este enfoque, le proporciona al proyecto la característica de ser escalable de forma que se consideren los futuros ciclos de mejora.

Los pasos que describen este marco de trabajo son seis: el primero es la comprensión del problema, donde se realiza un listado de requerimientos del proyecto entendiendo los *outputs* deseados como la generación de conocimiento relacionado a la madurez digital de las PyMEs Argentinas y el aprovechamiento de los programas de acompañamiento a la competitividad.[10]

El segundo paso es la comprensión de los *inputs*, y se inicia con la recolección y el estudio de la estructura y el origen de los datos. En este proyecto se definen como datos de entrada los formularios diagnósticos, los informes finales en formato PDF y los registros MiPyME en formato .CSV.

En tercer lugar, se preparan los datos de forma que el algoritmo desarrollado sea compatible con el formato de entrada de datos. En esta instancia se compatibilizan y generan algoritmos de soporte que permiten el manejo de errores en los datos de entrada.

Luego se modela el algoritmo según la técnica y elegida siendo en este caso una adaptación del *web scraping*[11] al formato de archivo PDF.

El quinto paso se considera una instancia de retroalimentación donde se evalúa si el algoritmo fue modelado correctamente o bien si la técnica elegida es adecuada y se valida con los objetivos iniciales.

Finalmente se establece la metodología de implementación y se puntean las recomendaciones para los siguientes ciclos de mejora.

2.2 Elección del algoritmo

El cuarto paso de la metodología de trabajo propuesta requiere de una previa investigación de los saberes y competencias del equipo de trabajo que presenta antecedentes de trabajos de aplicación de ciencia de datos para el abordaje de problemáticas que involucran datos abiertos a gran escala [12, 13]. Así como requiere también de una revisión de las técnicas apropiadas para llevar a cabo el modelado del algoritmo. Dicho esto, se elige el lenguaje de programación Python para la construcción del mismo y el entorno de desarrollo de código fuente desarrollado por Microsoft *Visual*

Studio Code (VSC). Ambas herramientas son gratuitas y de código abierto. En esta instancia se investiga sobre las técnicas actuales disponibles para la resolución del problema definido según los orígenes de datos por lo que, la técnica elegida es el *Web Scraping* y sus derivaciones en el *Scraping de PDF*.

El *scraping* se ha definido como un proceso automatizado cuyo objetivo es la extracción de información específica de la web. [11] Este concepto reemplaza a la búsqueda manual de información en una web para ser copiada y pegada en la base de datos del analista, por una búsqueda automatizada con scripts que utiliza parámetros específicos extraídos de los metadatos de la web.

Si bien la estructura puede variar según la información particular a recolectar y no existe un *framework* consensuado universalmente para realizar correctamente un *web scraping*, los pasos que se suelen seguir son los siguientes:

1) Entender el código fuente de las páginas web de donde se desea extraer los datos, para definir rutas de búsqueda.
2) Desarrollar un algoritmo que descargue las páginas y/o que extraiga los datos de las rutas asignadas.
3) Almacenar y procesar los datos obtenidos.

Esta metodología puede aplicarse de forma similar a documentos escritos en formato PDF sin necesidad de que estos estén disponibles en la web.

Para ello es necesaria la selección de la función de reconocimiento óptico del carácter o en inglés Optical Character Recognition (OCR) ya que es necesario el traspaso de formato del objeto reconocido de forma óptica al formato deseado.

En este trabajo la estructura principal que transforma el objeto reconocido del PDF a texto es ***pdfplumber.open(direction).extract_text()*** que usa la librería *pdfplumber* y las funciones dependientes *open* y *extract_text*.

Luego es necesaria la identificación de estructuras repetitivas y/o patrones en los documentos PDF que permitan la extracción de información. En este trabajo, se identificaron patrones basados en la cantidad y disposición de las páginas, así como la estructura de la información.

Posteriormente, se modeliza el algoritmo tal que reconozca diferentes patrones y se realizan numerosas validaciones de forma que siempre se extraiga la información requerida y que el proceso no se detenga ante un dato faltante.

Por último, se deben realizar una serie de operaciones de limpieza y normalización del texto extraído incluyendo en el algoritmo instancias que posibiliten el procesamiento de los documentos que no cumplan con los patrones establecidos, de forma de aumentar el porcentaje de procesamiento.

2.3 Modelado del algoritmo

Se estructura el modelo del algoritmo en función de los inputs y outputs que se resumen en la siguiente tabla:

Tabla 1. Inputs – Outputs del algoritmo.

INPUTS	Formato	OUTPUTS
Diagnósticos de los programas de acompañamiento: PAC 21 P.P. 4.0 22	PDF	Comparación del estado de madurez tecnológica y de mejora continua de las PyMEs acompañadas.
Informes finales de los programas de acompañamiento: PAC 21 P.P. 4.0 22	PDF	Exploración de representatividad de las PyMEs acompañadas respecto de los datos abiertos.
Registro MiPyME (Datos Abiertos 21)	CSV	Exploración de relaciones entre la mejora continua y la adopción de tecnologías 4.0

El objetivo general del algoritmo es lograr recolectar, analizar y compatibilizar los datos de interés de los documentos generados para los programas PAC 21 y P.P. 4.0 con las bases de datos abiertas del registro MiPyME, por lo que se dividen en dos conjuntos las funciones desarrolladas.

Por un lado, las relacionadas con el algoritmo para el *scraping* de PDF cuya finalidad es extraer la información de los diagnósticos e informes de los programas de acompañamiento: PAC 21 y P.P. 4.0 22 para transformarla en un archivo CSV que pueda ser fácilmente manipulado.

Por otro lado, el conjunto asociado a la comparación el análisis y visualización de las frecuencias de datos.

Como se puede observar en la Fig. 1 ambos conjuntos están formados por módulos que se relacionan entre sí. El conjunto de *scraping* de archivos en PDF está conformado por los módulos SCR, SCR2 y MAIN mientras que el conjunto de análisis de datos se conforma de dos módulos independientes: *Analisys* y *Wordcloud*.

Las relaciones entre conjuntos radican en el origen de datos que toma cada función que los compone y el objeto resultante que se retorna. Dicho esto, la ruta que sigue la información es la siguiente:

Primeramente, se indica la ruta de la carpeta principal que contiene las subcarpetas de cada una de las provincias que a su vez contienen los documentos por cada empresa de los programas. El modulo MAIN es quien recibe esta información y usa los módulos SCR 2 y SCR para devolver un archivo en formato CSV con la información solicitada y con los campos predeterminados.

Luego, se pueden utilizar diferentes parametrizaciones del algoritmo anterior para personalizar el archivo CSV según los datos de interés. Un ejemplo de esto son las parametrizaciones para extraer los niveles de madurez de los diagnósticos y la parametrización para extraer comentarios de los ATG.

Estos archivos son recibidos por los módulos *Analisys* y *Wordcloud* respectivamente y retornan gráficos de frecuencias de los datos analizados, pudiendo compatibilizar esta información porcentual con el mismo procesamiento en los registros nacionales.

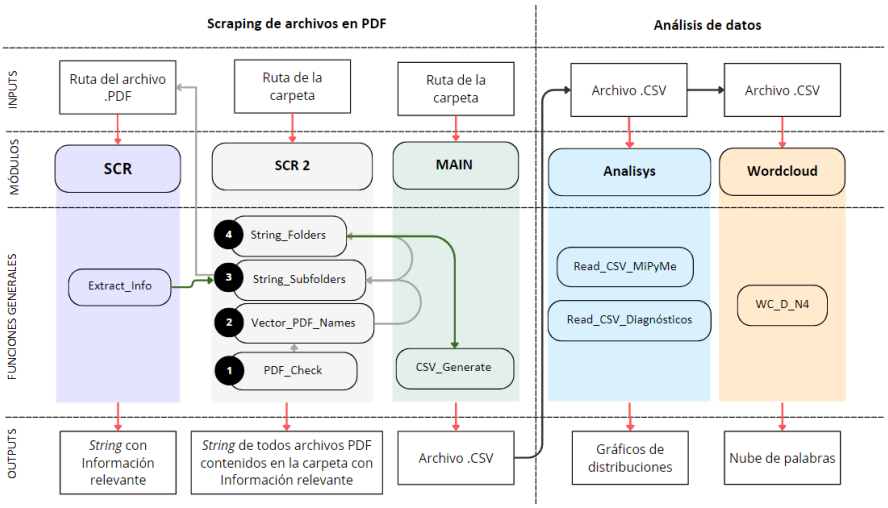

Fig. 1. Representación del modelo y sus relaciones.

A continuación, se explican los objetivos de cada uno de los módulos de forma específica y se desarrollan la concatenación de los algoritmos de los módulos principales.

Tabla 2. Objetivos de cada módulo del algoritmo *Scraping* de PDF

Nombre del Módulo	Objetivo
SCR	Extraer información específica de un documento PDF.
SCR2	Extraer la información específica de todos los documentos en PDF asignados en una carpeta.
MAIN	Devuelve un archivo en formato .CSV con la información específica de todos los documentos en PDF asignados en una carpeta dada su dirección.

Tabla 3. Descripción de los objetivos de cada módulo de los algoritmos de Análisis de Datos

Nombre del Módulo	Objetivo
Analisys	Retorna el listado de frecuencias de los campos requeridos en el CSV indicado e imprime los gráficos de sus frecuencias.
WordCloud	Muestra una nube de palabras que indica la frecuencia con la que se repiten los términos más significativos de lenguaje natural utilizado por los Asesores de Tecnologías de Gestión.

Tabla 4. Descripción del Pseudocódigo del módulo SCR

Pseudocódigo de SCR
Función Extract_Info (d) Intentar Abrir (d) PDF ← Archivo de la ruta “d” En caso de error Imprimir “Error”

```
        Definir T, P, NewString como cadena de caracteres
        Definir N como entero
        N ← Número de páginas de "d"
        Para  i ← 0 hasta N con paso 1 hacer
                P ← extraer texto de PDF
                T ← T + P
        Fin Para
        Intentar
                NewString ← NewString + extraer texto buscado de T
        En caso de error
                NewString ← NewString + "Error"
        // Comentario: La estructura que extrae los patrones del string "T" y los agrega a
        "NewString" se repite según la cantidad de patrones que se requieran extraer.  //

        Retornar NewString
Fin Función
```

Tabla 5. Descripción del Pseudocódigo del módulo SCR2

Pseudocódigo de SCR2

```
Importar SCR
Función PDF_Check(d)
        Si la extensión del archivo en d es igual a ".pdf"
        Retornar True
Fin Función

Función Vector_PDF_Names (s)
        Definir V, V1 como vector
        N ← Cantidad de archivos dentro de "s"
        V ← Archivos de la Subcarpeta "s"
        Para i ← 0 hasta N con paso 1 hacer
                Si Vi es un archivo Entonces
                        Si PDF_Check(Vi) Entonces
                                V1 ← Agregar al final Vi
      Fin Para
      Retornar V1
Fin Función

Función String_Subfolders (s, n)
        Definir V1 como vector
        V1 ← Vector_PDF_Names (s)
        Definir StringContainer como cadena de caracteres
        Definir L como entero
        L ← Largo de "V1"
       Si n es igual a -1 Entonces
        // Comentario: n= -1 indica que se procesarán todos los PDFs encontrados //
                Para i ← 0 hasta L con paso 1 hacer
                    StringContainer ← StringContainer + Extract_Info (s+V1i)
                Fin Para
        Retornar StringContainer
Fin Función
```

```
Función String_Folders (f, n)
        Definir StrFinal como cadena de caracteres
        N ← Cantidad de archivos dentro de "f"
        F ← Archivos de la carpeta "f"
        Para i ← 0 hasta N con paso 1 hacer
                StrFinal ← StrFinal + String_Subfolders (f +Fi, n)
        Fin Para
        Retornar StrFinal
Fin Función
```

Tabla 6. Descripción del Pseudocódigo del módulo Main.

Pseudocódigo de MAIN

```
Importar SCR 2
Procedimiento CSV_Generate (f, n= -1)
            // Comentario: "f" es una cadena de caracteres con la ruta de acceso a la
            carpeta principal //
            Definir CSV_string como cadena de caracteres
            CSV_string ← String_Folders (f, n)
            Intentar
                    Generar archivo ("Datos_INTI.csv", "w")
                    Escribir en "Datos_INTI.csv" (CSV_string)
                    Imprimir "Archivo creado exitosamente"
            En caso de error
                    Imprimir tipo de error
Fin Procedimiento
```

2.4 Resultados

Se trabajó con la comparación de los datos que ya se encontraban procesados del PAC 21 que conforman un set de 413 registros de cada una de las empresas participantes con las bases de datos abiertas (MiPyME) cuya cantidad de registros supera los 2 millones. Usando el módulo *Analisys* se procesan la totalidad de registros y se crean los gráficos de frecuencias en un promedio de 13,8 segundos en una computadora promedio de procesador i5 10th Gen (Intel(R) UHD Graphics), 8GB RAM, 256 SSD y OS: Windows 11, superando el procesamiento de la misma base de datos en Microsoft Excel que en las mismas condiciones sufre la pérdida de más del 50% de los registros.

A su vez, se utilizaron los tres módulos del algoritmo de *scraping* de PDF para incorporar al análisis comparativo los documentos del P.P. 4.0 22 almacenados en PDF. Se evaluó la efectividad del algoritmo con la totalidad de diagnósticos de las empresas participantes del P.P 4.0 2022 siendo la misma del 95,074%.

Se procesaron 203 documentos PDF de los cuales 10 presentaron formatos de imagen dentro de los mismos. El algoritmo toma un promedio de 80 segundos en recopilar toda información de los documentos y generar el archivo CSV con los campos de interés agregados en las mismas condiciones de hardware.

Gracias a esta comparación y al uso de modulo *Analisys* se lograron explorar las similitudes entre distribuciones provinciales y sectoriales de las empresas seleccionadas para cada Programa: PAC 21 y P.P. 4.0 22 respecto de los registros nacionales encontrando que las distribuciones provinciales porcentuales (ver Fig. 2.) se mantienen

en los programas demostrando y garantizando el acceso federal de todas las PyMEs en el país.

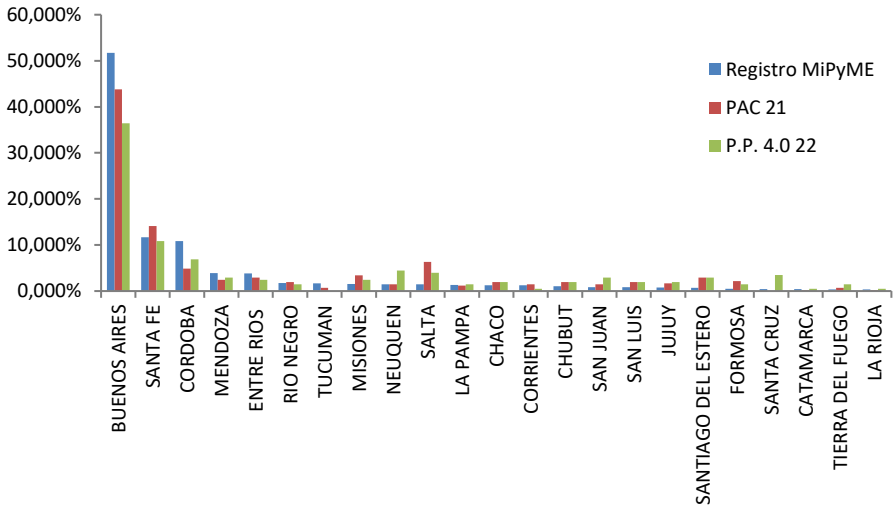

Fig. 2. Distribución porcentual provincial de las tres fuentes de datos evaluadas.

Luego, se contrastan los datos de ambos programas para comparar cuales son los niveles de *Kaizen* de las empresas participantes respecto de cada programa.
Como se puede observar en la Fig. 3. el nivel promedio de *Kaizen* para el programa de P.P 4.0 22 es más bajo que el del PAC 21 por lo que resulta interesante aclarar que algunas empresas pueden ver afectado su nivel de *Kaizen* debido a la incorporación de tres nuevos ítems que evalúan aspectos relacionados con digitalización para a incorporación de tecnologías 4.0 en el diagnóstico del P.P. 4.0.

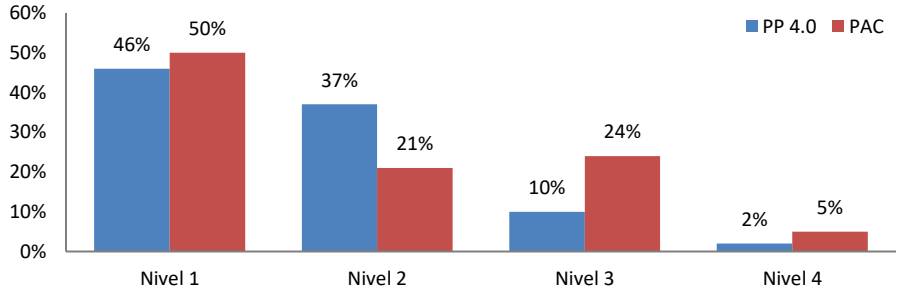

Fig. 3. Distribución porcentual del nivel de Kaizen en el PAC 21 y P.P. 4.0.

Finalmente, la extracción del texto escrito en los comentarios de los ATG que acompañaron estos programas permitió explorar cuales son las temáticas mayormente

mencionadas en el lenguaje natural de los mismos. A fines estratégicos, se decidió analizar únicamente los comentarios almacenados en los diagnósticos de las empresas cuyo nivel de *Kaizen* es 4, ya que se sabe desde la experiencia de los asesores que este nivel de madurez tiene una mayor tendencia a la digitalización. Dentro de los conceptos más frecuentes se encuentran, por ejemplo, la formulación de indicadores, el análisis de los registros y la gestión de la mejora continua como se puede observar en la Fig. 4.

Fig. 4. Nube de palabras de los comentarios de los ATG del P.P. 4.0 22 obtenido del modulo *Wordcloud* del algoritmo de Análisis de datos.

3 Discusión y líneas futuras

El proyecto de análisis de datos planteado inicialmente para la obtención de conocimiento relacionado al impacto de los programas de acompañamiento y la elaboración de herramientas que faciliten la asistencia sigue en proceso mostrándose exitosas las tres primeras etapas (ver Fig. 6.) relacionadas con la compatibilización de datos en formato PDF a formato CSV.

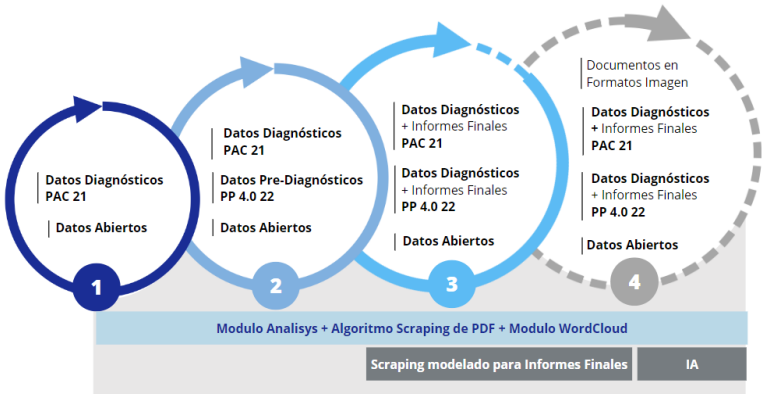

Fig. 6. Ciclos de mejora de los algoritmos en el transcurso del proyecto.

Los algoritmos desarrollados en Python han demostrado ser soluciones escalables y ágiles para el procesamiento de este tipo de documentos frente a otras herramientas

informáticas, sin embargo, se requieren conocimientos previos en el área de desarrollo de algoritmos.

La metodología de ciencia de datos elegida permite evaluar los siguientes ciclos de mejora del algoritmo identificando como requerimientos para el siguiente ciclo: la estandarización en la toma de datos en los registros de los programas futuros y la incorporación de sistemas de Inteligencia Artificial (IA) para el reconocimiento de textos en imágenes y de esta forma recuperar los archivos no procesables. (Ver Fig.6.)

Este modelo permitió el acceso a nueva información que hace posible el desarrollo de nuevas líneas de investigación ligadas al estudio de relaciones entre las empresas participantes y una muestra aleatoria que permita el análisis de representatividad del impacto de los financiamientos nacionales. Gracias a la escalabilidad y los potenciales modelados de patrones, el algoritmo permite plantearse las mismas hipótesis de forma tal que se pueda continuar entendiendo el comportamiento de las PyMEs argentinas frente a otros dispositivos de financiamiento para potenciar su competitividad.

References

1. Kulfas, M., Cafiero, S., Fernández, C., Fernández, A.: Plan de Dsearrollo Productivo Argentina 4.0. , Buenos Aires (2021).
2. Rodríguez, M., Rosso, J., Richard, A., Suárez, E., Pesci, R., Foti, S., Parenti, A., Zielinski, A.: Kaizen Tango: reflexiones de cómo ser más productivo en la Argentina. INTI, Buenos Aires (2019).
3. Strano, F., Gentile, N., Kunath Walsh, M.C., Lagier, M.E., Richard, A., Romanelli, M., Rosso, J.: Glosario TG: definiciones del entorno de la mejora continua. INTI, Buenos Aires (2022).
4. Carola, F., Korb, M., Lagier, M.E., Richard, A.: Guía TG: Metodología de intervención de la Red de Tecnologías de Gestión en PyMEs. INTI, Buenos Aires (2020).
5. Pascal, G., Tornillo, J.E., Torres, Z., Redchuk, A.: Implementación de Sistemas de Soporte de Decisión en Universidades: La Estructuración y Modelización de Problemas de Objetivos Múltiples, http://www.clei2017-46jaiio.sadio.org.ar/sites/default/files/Mem/SIE/SIE-14.pdf, (2017).
6. Walas, F., Tornillo, J.E., Orellana, V., Fretes, S., Seminario, A.G.: Analysis of the approach in Local SMEs to Production 4.0 tools View project. (2022).
7. Ministerio de Desarrollo Productivo: Datos Argentina - Registro MiPyME, https://datos.gob.ar/dataset/produccion-registro-mipyme, last accessed 2023/05/05.
8. Martínez, R., Vilaboa, P., Catala, N.: Evaluación de algoritmos de aprendizaje con datos públicos abiertos de machine learning mediante Orange3. Memorias de las JAIIO. 8, 58–68 (2022).
9. Schröer, C., Kruse, F., Gómez, J.M.: A systematic literature review on applying CRISP-DM process model. Procedia Comput Sci. 181, 526–534 (2021). https://doi.org/10.1016/J.PROCS.2021.01.199.
10. Ministerio de Industria y Desarrollo Productivo.: Programa de Apoyo a la Competitividad (PAC) , https://www.argentina.gob.ar/produccion/programa-de-apoyo-la-competitividad-pac, last accessed 2023/05/10.
11. Marrero, L., Olsowy, V., Thomas, P.J., Delía, L.N., Tesone, F., Fernández Sosa, J., Pesado, P.M.: Análisis de técnicas de raspado de datos en la web aplicado al Portal del Estado Nacional Argentino. XXV Congreso Argentino de Ciencias de la Computación (CACIC) . 1–9 (2019).

12. No, I.N., Tornillo, J.E., Pascal, G., Rabbione, L.: Ciencia de Datos y Reportes de Movilidad Google para Modelizar la Demanda de Combustible. Memorias de las JAIIO. 8, 20–24 (2022).
13. Pascal, G., Tornillo, J.E., Minnaard, C., Comoglio, M.: Data mining to increase teaching performance in engineering education. ACM International Conference Proceeding Series. 308–311 (2019). https://doi.org/10.1145/3318396.3318433.