

DEL PERIODISMO DE DATOS A LA INTELIGENCIA ARTIFICIAL: DESAFÍOS QUE ENFRENTA LA NACIÓN EN LA IMPLEMENTACIÓN DE LA VISIÓN POR COMPUTADORA EN LOS REPORTAJES DE NOTICIAS*

Estudiante:

Ginger Valuarte

October 2023

*Matías Felipe de Lima Santos*¹:

Ramón Salaverría²

*Recibido: 10/09/2020 Aprobado
por pares: 11/08/2021*

*Enviado a pares: 13/01/2021 Aceptado:
09/09/2021.*

DOI: 10.5294/pacla.2021.24.3.7

Para hacer referencia a este artículo / para citar este artículo / para citar este artículo
de-Lima-Santos, MF., Salaverría, R. (2021). Del periodismo de datos a la inteligencia artificial: desafíos que enfrenta La Nación en la implementación de la visión por computadora en los informes de noticias. Palabra clave, 24(3), e2437. DOI: <https://doi.org/10.5294/pacla.2021.24.3.7>

Abstract

El periodismo está en un punto de cambio radical, impulsado por el aumento de datos, sensores y avances tecnológicos. Aunque la IA ha aportado beneficios significativos a varias áreas del periodismo, el área de la investigación es uno de los pocos que no lo implementan del y sigue siendo un desafío debido a su novedad y complejidad. El estudio se centra en los desafíos de implementar proyectos de noticias con visión por computadora en “La Nación”, un periódico líder en América Latina. Se identifican cuatro dificultades principales a la hora de implementar la IA como son: la falta de imágenes de alta resolución, la falta de infraestructura tecnológica disponible, la ausencia de personal calificado para desarrollar los códigos necesarios y un proceso de implementación largo y costoso. Este artículo concluye

con una discusión sobre la centralidad de las soluciones de IA en manos de las grandes corporaciones tecnológicas.

1. Introducción.

Según Knight (2015) y Lewis y Usher (2013), la ubicuidad de las tecnologías de la información y las comunicaciones (TIC) en las noticias y la datificación de la sociedad han aumentado las oportunidades para utilizar nuevos formatos para crear contenidos. Los periodistas han descubierto un siguiente paso lógico a través de un mejor acceso a datos y metadatos. Es posible procesar datos para producir información útil que ayudará en el análisis y presentación de informes sobre cuestiones de interés público. Al principio, el periodismo de datos estuvo fuertemente influenciado por Occidente, y la mayoría de las organizaciones de noticias de élite procedían del hemisferio norte, incluidos Estados Unidos, el Reino Unido, las naciones nórdicas y Alemania (Anderson, 2019; Appelgren Nygren, 2014; Borges-Rey, 2016; Coddington, 2015; De Maeyer et al.). Karlsen y Stavelin, 2014; Lewis y Westlund, 2015; Felle, 2016; Fink y Anderson, 2015; Fink y Anderson; Fink y Anderson; Stalph). En los últimos años, el creciente interés de los académicos por el periodismo de datos ha estimulado un importante cuerpo de literatura sobre el desarrollo de la presentación de datos en el Sur Global (Mutsvairo et al., 2019). En el mito del periodismo de datos, los profesionales recibieron poder gracias a la capacidad computacional superior del siglo XXI porque les permitió pasar de la era del big data a la era de la inteligencia artificial (IA), cuando los principios de la informática y la ciencia de los datos se volvieron más y más común en el periodismo (Borges-Rey, 2021). El diario argentino La Nación es un ejemplo de ello (Mazotte, 2017; Palomo et al.); sin embargo, pocos casos fuera del mundo occidental han tenido éxito y han sido reconocidos tanto en la comunidad periodística como en la académica. , 2019). A pesar de que la mayor parte del Sur Global, especialmente América Latina, carece de datos estadísticos fácilmente accesibles (Salaverra, 2016, p. XVI), La Nación ha podido involucrar a su audiencia e innovar a través de la difusión de noticias. El equipo de periodismo de datos de los medios ha logrado reconocimiento tanto interno como externo como resultado del involucramiento exitoso de la audiencia en el proceso informativo (Palomo et al., 2019). eISSN: 2027-534X; ISSN: 0122-8285; vol. 24 No. 3. Además de los datos abiertos, el periodismo cívico y público (Lambeth, 1998) y el periodismo automatizado (Carlson, 2016), el equipo detrás de los proyectos de periodismo de datos, Data de La Nación (Datos LN), tiene experiencia en todos de estas áreas para mejorar las experiencias del lector y agregar valor para la audiencia. Como resultado, las organizaciones de noticias deben idear conceptos y formatos nuevos (Franklin, 2014; Picard, 2014). Data LN ha buscado continuamente nuevas

formas de interactuar e informar a su audiencia para prever el futuro de la industria. La Nación sigue limitada por el hecho de que es una organización de medios tradicional ubicada en una nación que frecuentemente se ve sacudida por crisis políticas y económicas, a pesar de ser elogiada por sus habilidades innovadoras en materia de datos. Los recursos son limitados y la innovación es gradual en los medios latinoamericanos, como lo es en muchos otros medios (Salaverra y de-Lima-Santos, 2021). También es importante tener en cuenta que, en los mercados tecnológicos más desarrollados, la innovación incremental frecuentemente triunfa sobre el cambio radical, lo que lleva a mejores procesos y productos (Eisenman, 2013; Verganti, 2006). LNDato ha estudiado cómo aplicar técnicas de IA en la producción periodística en un esfuerzo por romper esta burbuja de innovación incremental en las redacciones (Paulussen, 2016). Mediante el uso de algoritmos, la inteligencia artificial (IA) permite a las máquinas imitar el comportamiento humano y aprender de sus experiencias, así como adaptarse a nuevas entradas. La visión por computadora (CV), una de las muchas aplicaciones de la IA, se ocupa principalmente del procesamiento de imágenes. Al contrario de lo que su nombre podría implicar, los modelos CV son incapaces de percibir el contenido de una imagen como lo hace el ojo humano (Marr, 2010). En cambio, infieren el contenido mostrado utilizando algoritmos matemáticos (Szeliski, 2011). El proceso de edición se puede acelerar utilizando CV, que también puede "permitir a los periodistas obtener pruebas para artículos de investigación.^{al} clasificar y organizar rápidamente una gran cantidad de imágenes y vídeos (Marconi et al., 2017, p. 14). Con esto en mente, Data LN desarrolló un proyecto basado en CV para mapear los parques solares de rápido crecimiento de Argentina que se adhieren al Acuerdo de París para reducir su huella de carbono y adoptar energía 100 por ciento renovable para controlar el aumento de la electricidad. (Masson-Delmotte et al.) informan una temperatura promedio global de menos de 2 °C. , 2018). Exploramos los desafíos que implica la realización de un proyecto de CV basándose en la literatura sobre gestión, periodismo y teorías de la informática. 6 Del periodismo de datos a la inteligencia artificial. que se está desarrollando rápidamente en el campo del aprendizaje automático, incluido Mathias Felipe de Lima Santos y otros. Esta área aún está poco investigada en comparación con otros usos de la IA en el periodismo porque la mayor parte de la literatura analiza cómo el periodismo automatizado afecta la información periodística (Salaverra y de-Lima-Santos, 2020), lo que frecuentemente implica un modelo que completa los espacios en blanco de Historias modelo preescritas sin un modelo de aprendizaje automático (Biswal yGouda, 2020). Nos basamos en un experimento realizado por La Nación entre enero y marzo de 2020 para comprender este fenómeno. De acuerdo con los privilegios del investigador dentro de la organización bajo investigación, los métodos de investigación cualitativa tienen el potencial de extraer conocimiento de manera efectiva. La observación participante le da al investigador la oportunidad de interactuar con miembros del

personal de todos los rangos durante un estudio etnográfico. El primer autor pudo trabajar junto al equipo de datos de The Nation en su oficina en Buenos Aires, Argentina, como resultado de lo cual llevamos a cabo la observación participante para este estudio. El Luego, los datos se triangularon con información adicional de las entrevistas, validando la información recopilada y permitiéndonos buscar inconsistencias en los informes (O'Donoghue, 2003). Debido a que en este estudio se utilizaron la observación participante y las entrevistas en conjunto, se utilizó un enfoque de métodos múltiples. Este estudio etnográfico que realizamos analizó las dificultades que tiene el departamento de datos de The Nation para integrar nuevos conocimientos tecnológicos en la creación de historias de datos. Nuestro objetivo final fue dar respuesta a la siguiente pregunta de investigación: ¿Qué dificultades encuentra LN Data al implementar CV, un tipo de IA, con fines periodísticos? el periodismo no sólo avanza el conocimiento académico del tema sino que también proporciona conocimientos prácticos que pueden ayudar a los periodistas a perfeccionar su oficio. En segundo lugar, al resaltar nuevos problemas y amenazas a la implementación de proyectos tecnológicos de vanguardia y concentrarse en organizaciones del Sur Global, como La Nación, este artículo contribuye al trabajo académico del periodismo. Finalmente, este trabajo presenta - Palabra clave - ISSN: 0122-8285 - eISSN: 2027-534X - Vol. 3 posibles dificultades que pueden encontrar las organizaciones de noticias al implementar proyectos de IA en sus redacciones. las etapas reales de implementación de proyectos de imágenes satelitales, CV o análisis de imágenes en periodismo.

2. Revisión de literatura.

Transformación: aportando capacidades dinámicas y nuevas estrategias corporativas

Las tecnologías digitales han catalizado cambios significativos y perturbaciones del mercado en la industria de los medios (Oliver, 2018; Teece, 2010). En esta época de cambios complejos, la mayoría de las empresas de medios no logran sacar provecho de los productos digitales, es por eso que las organizaciones de medios deben adoptar estrategias y adaptarse a este cambio tecnológico masivo y confiar en una serie de procesos que involucran la tecnología en la producción de noticias para lograr sus objetivos. En el periodismo, este proceso suele ser conservador y lento (Kueng, 2017; Lischka, 2019) debido al síndrome del “sí, pero” (YBS), que limita las nuevas estrategias digitales por desafíos económicos y otras excusas como la falta de recursos. y personal, la viabilidad económica o el riesgo de fracaso (Paulussen, 2016). La innovación no es una acción temporal sino una estrategia a largo plazo pues implica nuevas formas de involucrar a audiencias establecidas y maduras más allá del determinismo

tecnológico debido a que permite incluir nuevas formas de informar y narrar historias, como blogs, periodismo ciudadano, realidad virtual y aumentada, datos y colaboración, y aportar rentabilidad a los modelos de negocio digitales.

Periodismo de datos: desafíos y oportunidades en el Sur Global.

El origen del periodismo de datos tiene sus raíces en el reportaje asistido por ordenador (CAR), que data de 1973 y se basa en el trabajo de Philip Meyer (2002). Al combinar encuestas, análisis de contenido y estadísticas, Meyer validó las preguntas de los periodistas y adoptó una nueva forma de reportaje periodístico (McGregor, 2013). Hasta hace poco, se sabía poco sobre el desarrollo del periodismo de datos más allá de las sociedades occidentales; Gran parte de la literatura académica y los casos prácticos se concentraron en países pioneros como Estados Unidos, el Reino Unido y los países nórdicos (Anderson, 2019; Appelgren Nygren, 2014; Borges-Rey, 2016; Felle, 2016; Fink Anderson, 2015; Karlsson y Stavelin, 2014; Lewis y Westlund, 2015; Stalph, 2018).

Requirió un nuevo nivel de alfabetización, habilidades de escritura y aritmética, conocimiento de la tecnología y una mentalidad analítica, lo que podría explicar el dominio de Estados Unidos, el Reino Unido y los países nórdicos en el periodismo de datos. Estas naciones históricamente han enfatizado la transparencia, el acceso a la información y la digitalización.

A pesar de estas limitaciones, la ArgentinaLa Naciónha podido descubrir historias de datos significativas. El antiguo medio de noticias se fundó en 1870 y se convirtió en un ejemplo de éxito en la narración de datos. Su equipo dedicado,Datos LN, ha recibido un importante reconocimiento por sus esfuerzos por superar la falta de datos y recursos para contar historias.Un ejemplo de sus esfuerzos es el hackathon cívico durante el cualDatos LNsolicitó voluntarios del público para ayudar a analizar 4.800 documentos públicos verificados relacionados con los gastos del Senado entre 2010 y 2012 (Palomo et al., 2019).En este sentido, el equipo desarrolla múltiples soluciones alternativas basadas en la comprensión de los requisitos críticos del equipo y la sala de redacción,incluido el diseño de scripts y algoritmos para automatizar tareas y resolverproblemas computacionales y el despliegue de soluciones internas pararespaldar el trabajo de datos y equipos infográficos. Esto concuerda con los hallazgos de estudios que sugieren que muchos experimentan cierto grado de frustración con las soluciones genéricas de terceros para garantizar que su oferta sea inclusiva y lo suficientemente amplia para respaldar la producción periodística.

Inteligencia artificial: una industria en auge que el periodismo debe explorar en profundidad.

El primer uso del término “inteligencia artificial” se remonta a 1955, cuando John McCarthy, profesor de la Universidad de Stanford, describió la ciencia y la ingeniería de la fabricación de máquinas inteligentes (McCarthy, 1998). Desde entonces, el término se ha extendido ampliamente y actualmente no existe una definición universal de IA, puede describirse simplemente como el proceso de automatización de tareas cognitivas (Chan-Olmsted, 2019) o, en otras palabras, como un área de conocimiento “dedicada a la creación de máquinas y sistemas informáticos que realicen operaciones análogas al aprendizaje y la toma de decisiones (Castro New, 2016, p. 2).

La ubicuidad de la informática se ha hecho evidente y ha aportado beneficios incommensurables a todos los campos de estudio, especialmente a la ciencia. Sin embargo, según algunos autores, la IA es un concepto controvertido que aún no se ha plasmado plenamente en el periodismo; a lo mejor, lo que existe actualmente es una simple automatización que llena los espacios en blanco de las historias modelo (Biswal Gouda, 2020).

Dada la relativa novedad y complejidad del tema, pocas organizaciones han logrado implementar soluciones personalizadas para sus negocios, solo las grandes empresas tecnológicas como Facebook, Amazon, Apple, Netflix y Google aprovechan su control sobre la IA adquiriendo activamente empresas desconocidas que desarrollan soluciones de IA para concentrar más poder (Linden, 2017a, 2017b). La IA está comenzando a abrirse camino en una amplia gama de proyectos en la industria de los medios, desde servicios hasta productos, incluido, por ejemplo, software de IA avanzado que automatiza la verificación de datos mediante el análisis de grandes cantidades de datos.

Sin embargo, la IA no se limita al contenido textual. Recientemente, los investigadores han encontrado áreas prometedoras que la IA puede potenciar. Por ejemplo, Diakopoulos (2015) presentó desafíos y oportunidades relacionados con la utilización del periodismo automatizado en las redacciones, mientras que Graefe (2016) presentó la perspectiva de la audiencia sobre los textos automatizados. Linden (2017b) estudió la respuesta de los medios al contenido automatizado en Estados Unidos y otros cinco países europeos. Estos avances en las tecnologías relacionadas con la IA tienen el potencial de alterar significativamente la naturaleza de las interacciones entre humanos y máquinas.

Sin embargo, aún falta una comprensión clara de las oportunidades y riesgos de la IA a corto y largo plazo. Los mecanismos de aplicación de la IA son esenciales para garantizar que los sistemas de IA cumplan con las pautas legales y éticas sin considerar explícitamente las estructuras de poder entre las distintas partes interesadas, por este motivo, es fundamental comprender los diferentes dominios y subconjuntos de la IA.

Visión por ordenador, una rama de la inteligencia artificial

La IA utiliza algoritmos (un procedimiento paso a paso para resolver problemas) para permitir que las máquinas aprendan de las experiencias, se ajusten a nuevas entradas y realicen tareas similares a las humanas. Hasta hace poco, la IA solo funcionaba con una capacidad limitada, ya que los informáticos tenían que programar una amplia gama de funciones en un sistema para imitar la inteligencia humana. En los últimos años, esta capacidad de IA ha aparecido debido a un mejor hardware, más datos, mejores algoritmos. El aprendizaje automático es una herramienta poderosa que permite a los humanos ver cosas que de otro modo no podrían ver cuando los conjuntos de datos superan un cierto tamaño” (Marconi, 2020); en consecuencia, el ML es un subconjunto de la IA que realiza una tarea a través de un sistema sin instrucciones explícitas o con una asistencia mínima de los programadores.

El desarrollo del ML está muy relacionado con el aprendizaje profundo, dependiendo del nivel de actividad humana, el aprendizaje profundo se puede clasificar en supervisado, semisupervisado o no supervisado (Castro y New, 2016; Marconi, 2020). El aprendizaje supervisado implica una tarea de aprendizaje automático entrenada para asociar un objeto con el resultado deseado. Por lo tanto, las técnicas basadas en aprendizaje supervisado requieren una gran cantidad de datos de entrenamiento anotados por humanos para aprender un modelo adecuado, y es común en problemas que involucran clasificación de imágenes, subtítulos de imágenes, segmentación de instancias, respuesta visual a preguntas y otras tareas.

Por el contrario, el ML no supervisado es libre de generar relaciones entre entrada y salida sin un resultado objetivo. Habitualmente, esta tarea se realiza comparando similitudes y desviaciones entre la información encontrada en un conjunto de datos (Marconi, 2020). Los algoritmos de aprendizaje semisupervisados representan un término medio entre los métodos supervisados y no supervisados. El aprendizaje semisupervisado etiqueta una pequeña cantidad de datos y deja otra gran parte sin etiquetar durante el entrenamiento. Al encontrarse entre estos dos dominios, el aprendizaje semisupervisado requiere menos interacción humana que el aprendizaje automático no supervisado (Aronson, 2018; Marconi, 2020).

En periodismo, el CV puede acelerar el proceso de edición y permitir a los periodistas obtener pruebas para artículos de investigación. Hasta ahora, los académicos han prestado poca atención al CV y sus implicaciones para el periodismo (Chan-Olmsted, 2019).

3. Método.

Se basa en una estrategia de investigación cualitativa utilizando un diseño multimétodo que utilizó observación participante y entrevistas en profundidad, esta investigación tiene como objetivo propor-

cionar una visión más holística y matizada de los desafíos que enfrentan las redacciones al adoptar la IA, especialmente el CV.

La observación en la sala de redacción consistió en aproximadamente tres meses de trabajo de campo en el primer trimestre de 2020, antes de que la pandemia golpeará a América Latina, cuando el primer autor estaba en comisión de servicio en la organización. En promedio, el primer autor pasó 40 horas por semana con el equipo. La observación incluyó notas cualitativas que el primer autor registró durante o después de cada día en la sala de redacción. Además, el primer autor participó en reuniones que habitualmente duraban 60 minutos. Durante estas reuniones y el período pasado en la institución anfitriona, varias personas y organizaciones internas y externas se movilaron en torno a estos temas, lo que permitió al primer autor entablar mayores interacciones con ellos. Estos actores en las reuniones ofrecieron otras perspectivas sobre temas ambientales y tecnológicos que fueron útiles para implementar este nuevo proyecto.

Los datos obtenidos fueron analizados utilizando el método propuesto por Emerson et al. (2011), que consta de un proceso de tres pasos. Primero, el primer autor examinó de cerca y sistemáticamente las notas de campo para encontrar patrones. En segundo lugar, los autores realizaron una codificación analítica cualitativa en dos fases: codificación abierta que permitió al etnógrafo leer las notas de campo línea por línea y formular todos los temas e ideas en ellas y, luego, codificación enfocada en la que sometió las notas a un análisis detallado que utilizó un conjunto más pequeño de ideas y categorías basadas en esos temas. En la última fase, escribimos nuestros hallazgos como “cuentos” (Richardson, 1990). A diferencia de los investigadores cuantitativos que expresan sus hallazgos en tablas y gráficos, los investigadores cualitativos deben depender de las palabras. Los datos de observación se triangularon con datos auxiliares de las entrevistas en profundidad. Al escuchar los relatos de los entrevistados y tomar notas de los eventos importantes en las narrativas, generamos conocimientos confiables sobre el fenómeno bajo investigación, lo que nos permitió ir más allá de nuestras observaciones iniciales para complementar la información faltante (Becker y Geer, 1957; Bryman, 2012; Kawulich, 2005).

3.1. Inspiración.

Es primordial describir el trabajo preliminar realizado antes de que esta idea llegara a la redacción para explicar por qué La Nación decidió experimentar con CV en la producción de noticias. Mencionar al equipo profesionales que provienen de una amplia gama de procedencias que estuvo formado por: un periodista, un tecnólogo, un abogado y un bibliotecario. Angélica “Momi” Peralta, tecnóloga, lidera el equipo desde su fundación. Sus valores se reflejan en la mentalidad multicultural del equipo. Sus

diversos orígenes son una de las razones más importantes para lograr resultados extraordinarios.

Florencia Coelho desempeña un papel fundamental en el equipo como Gerente de Capacitación e Investigación de Nuevos Medios y es responsable de encontrar nuevas ideas y oportunidades que se adapten a los objetivos del equipo. Su trabajo también consiste en seguir las tendencias internacionales y entender qué podría ser útil para su redacción. Con esto en mente, entendió que el equipo necesitaba comenzar a experimentar con niveles más avanzados de aplicación de tecnología para la presentación de noticias, principalmente desde una perspectiva de inteligencia artificial. Sin embargo, como relató en una entrevista, “la IA me pareció un tema muy interesante, pero no entendí nada antes de verlo por primera vez” (Florence Coelho, Datos LN, enero de 2020). Por lo tanto, en 2019, se convirtió en JSK Fellow en Stanford para dedicar plenamente su tiempo a aprender sobre las oportunidades para aplicar la IA al ecosistema de noticias.

Mientras tanto, Datos LN se inspiró para informar sobre la emergencia climática después de colaborar con Covering Climate Now de Revista Columbia.

proyecto impulsado por Fundación Avina. Para estos proyectos, el equipo ha cubierto noticias medioambientales mundiales como los incendios en Australia y la selva amazónica y las charlas de Greta Thunberg en reuniones internacionales (Cumbre de las Naciones Unidas, COP 25), al tiempo que intenta aprovechar las preocupaciones locales sobre cuestiones medioambientales. Impulsada por estas preocupaciones sobre el cambio climático, la industria de las energías renovables en todo el mundo ha experimentado un crecimiento significativo. A principios de 2016, durante el mandato del expresidente Mauricio Macri, Argentina lanzó un ambicioso programa llamado Renovar, que prometió aprovechar los abundantes recursos de energía limpia del país promoviendo la generación privada de energía renovable a través de una subasta. Renovar, acrónimo de Subastas de Energías Renovables de Argentina, se esforzó por aumentar las fuentes de energía renovables al 20 por ciento de la matriz energética nacional para 2025 (IFC, 2018).

El proyecto garantiza al país algunos puntos del Acuerdo de París ratificado, el acuerdo histórico firmado en 2016 que reunió a 189 partes en todo el mundo para combatir el cambio climático. Según los hallazgos de múltiples estudios de la Agencia Internacional de Energías Renovables (IRENA), la energía renovable tiene el potencial de cumplir el 90 por ciento de los objetivos de reducción de emisiones relacionadas con la energía del Acuerdo de París (CMNUCC, 2017). Argentina posee vastos recursos naturales para crear energía limpia, especialmente energía eólica y solar, lo que le permite al país avanzar hacia la energía renovable.

El sector fue supervisado por el Ministerio de Energía y Minería (MI-NEM [Ministerio de Energía

y Minería]), que era el encargado de definir las políticas para el sector energético y supervisar su aplicación. Renovarse dividió en varias rondas de subasta, que establecieron el proceso para atraer postores internacionales y crear un nuevo mercado para la inversión privada en energías renovables. A finales de 2016, Argentina había realizado con éxito dos rondas de subastas. Otras rondas hicieron de este un proyecto exitoso. Sin embargo, aceptar una oferta en una subasta no garantizaba su entrega en una fecha determinada ni la posibilidad de cambios en el proyecto. Una de las mayores inspiraciones provino del proyecto Deep Solar de Stanford. El equipo que trabajó en este proyecto utilizó un modelo de aprendizaje profundo para detectar paneles solares en imágenes satelitales (ver Yu et al., 2018). De este modo, Datos LN decidió utilizar un modelo predictivo basado en aprendizaje automático para estimar el despliegue de parques solares en el país y, al mismo tiempo, probar y aplicar nuevas tecnologías a la producción de noticias. Nuestros hallazgos se describirán en las siguientes secciones.

Nuevas tecnologías en el periodismo: un proceso de aprendizaje constante

El uso de nuevas tecnologías en el periodismo plantea preguntas novedosas para la comunidad sobre cómo informar mejor sobre los problemas, convertir los datos en conocimientos y contar historias convincentes que impacten a la audiencia. En el Renovar proyecto, era necesario un subconjunto de IA y ML para desarrollar el proyecto. CV ayudó a detectar la ubicación de los parques solares e identificar el estado de los proyectos para comprobar si los parques se implementaron dentro del plazo propuesto por el gobierno.

Sin embargo, durante el proceso, Datos LN enfrentó algunos problemas mientras desarrollaba el experimento para crear prototipos de soluciones impulsadas por IA utilizando imágenes satelitales. Entre los desafíos clave, “el equipo no tenía la experiencia para implementar algoritmos de visión por computadora” (Florencia Coelho, Datos LN, enero de 2020). Su primer paso fue buscar otros proyectos que adoptaran enfoques similares. Un par de estudios abordaron el uso de imágenes satelitales para reconocer paneles solares, como el proyecto de techo corredizo de Google, el mapeo solar fotovoltaico de OpenClimate-Fix, Deep Solar de Stanford en los EE. UU., un estudio polaco de la Universidad de Economía y Negocios de Poznań, y Techo corredizo de India. creado por una startup. Al analizar estos estudios, el equipo entendió que la complejidad de la tarea aumentaba debido a la calidad de las imágenes satelitales en el mundo en desarrollo, incluida Argentina. La falta de una unidad de investigación en el espacio como la NASA en Estados Unidos, la Agencia Espacial Europea (ESA) o el INPE (Instituto Nacional de Investigaciones Espaciales) en Brasil complicó aún más el acceso a imágenes de alta calidad.

Imágenes de satélite: un tesoro precioso, pero restringido a un número mínimo de

actores

El primer desafío, identificable a nivel local y nacional, implicó comprender las imágenes necesarias para alimentar el algoritmo y el nivel de precisión deseado. Un algoritmo CV busca comprender y automatizar tareas que el sistema visual humano puede realizar, lo cual sólo es posible proporcionando imágenes visibles para el ojo humano. Esto se puede medir utilizando la resolución espacial, que se refiere al tamaño de píxeles de una imagen de satélite, la unidad básica de color programable en una imagen de computadora.

La mayoría de las opciones de imágenes satelitales son costosas; adquirirlos no era realista para La Nación. Sentinel-2, que comprende satélites gemelos que vuelan en la misma órbita pero con una fase de 180°, forma parte del programa de observación de la Tierra de la Unión Europea coordinado y gestionado por la Comisión Europea. Debido a sus datos gratuitos disponibles con una resolución espacial relativamente alta y un alto tiempo de revisita (tiempo de revisita de 5 días), Sentinel-2 tenía el mayor potencial para proporcionar las imágenes necesarias para identificar parques solares. Estos satélites gemelos son una de las pocas opciones gratuitas y con una política abierta que permite su uso con este fin.

Aunque el tamaño de píxel era de alcance moderado en comparación con otras opciones disponibles en el mercado (que no eran gratuitas), los parques solares dejan una huella grande y un patrón reconocible, ya que requieren una cantidad significativa de energía solar fotovoltaica. Las imágenes de satélite geoestacionarios también resultan útiles para grandes áreas, principalmente países con grandes extensiones territoriales, como Argentina, Brasil y Estados Unidos. Otras soluciones, como Planet, ofrecen imágenes multiespectrales mucho mejores con resoluciones más altas, pero el acceso a ellas es costoso para una sala de redacción del Sur Global. Como se mencionó anteriormente, otro tema importante fue la frecuencia de actualización. Se mapeó un proyecto en construcción mientras se recogían los datos; por lo tanto, eran necesarias imágenes de un satélite diseñado para proporcionar una alta frecuencia de revisita. Aunque Google ofrece imágenes de alta calidad a través de su API, como se ilustra en la Figura 1, no tiene un cronograma fijo de actualizaciones y la calidad de la imagen varía según el área.

FIGURA 1. PARQUES SOLARES EN NONOGASTA, LA RIOJA, ARGENTINA (-29.328923, -67.423753). FUENTE: IMÁGENES CAPTURADAS DE GOOGLE MAPS,

SENTINEL-2 Y PLANET (PLANET, 2020; SENTINEL-2, 2020).



Figura 1: mapas de Google



Figura 2: Centinela-2



Figura 3: Planeta

En este contexto, el equipo decidió que Sentinel-2 era su mejor opción porque está diseñado para proporcionar una frecuencia de revisita de cinco días en el Ecuador, similar a servicios comerciales como Planet pero con mejor resolución. Hay que tener cuidado con lo que utilizan para mapear imágenes de satélite. 22 Del Periodismo de Datos a la Inteligencia Artificial... - Mathias Felipe de Lima Santos y otros La resolución, las similitudes y las incertidumbres plantean un riesgo para la información periodística, como se refleja en otros estudios (Marconi et al., 2017).

La prerrogativa se limita a las empresas de tecnología: infraestructura y poder informático masivo

La prerrogativa se limita a las empresas de tecnología: infraestructura y poder informático masivo El objetivo era explorar el desarrollo de RenovAr en todo el país, lo que corresponde a una superficie de 2.780.000 km². Representa una enorme extensión territorial ya que Argentina es el octavo país más grande del mundo. Como tal, la potencia informática necesaria para analizar las imágenes descargadas de Sentinel-2 requería hardware avanzado y recursos de computación en la nube, que Datos LN no tenía. El principal beneficio de utilizar la computación en la nube es tener acceso a potencia e infraestructura informática que sería difícil y costoso establecer individualmente; sin embargo, la computación en la nube sigue siendo costosa y limitada a las grandes empresas tecnológicas, como la IA (Chan-Olmsted, 2019).

Luego de reuniones consecutivas con diferentes organizaciones, el Datos LN el equipo se asoció con una nueva empresa especializada en aplicaciones de datos satelitales que hacen que varios objetos sean detectables a través de CV. La nueva empresa argentina Dymaxion Labs ofreció una API GeoML basada en la nube para extraer información sobre observación de la Tierra y proporcionó la infraes-

estructura que no se encuentra en la sala de redacción. Fue una situación beneficiosa para ambos: La Nación podría usar sus recursos para producir más conocimiento en un área en la que el equipo aún no tenía experiencia, y Dymaxion Labs podría aplicar su software a un proyecto de servicio público.

Usando visión por computadora para mapear parques solares argentinos

En este caso, MINE proporcionó una lista de lugares donde se construirían granjas solares. Con base en estos datos, se creó una lista de lugares que ya tenían parques solares, pero la cantidad de lugares con parques solares construidos en diciembre de 2019 no era lo suficientemente alta como para generar un conjunto de datos de capacitación. Para solucionar esto, los equipos descubrieron geolocalizaciones en OpenStreetMap de los parques solares de Chile. Al unir estos dos conjuntos de datos, se creó un conjunto de datos final. Ambos equipos analizaron una muestra de 100 geocoordenadas utilizadas para entrenar el conjunto de datos. Estos parques solares fueron revisados manualmente y clasificados. Resumidamente, los pasos del algoritmo fueron los siguientes:

1. Se dibujaron polígonos alrededor de granjas solares conocidas según el conjunto de datos construido;
2. Se recortaron contornos geolocalizados de granjas solares conocidas a partir de imágenes satelitales desde junio de 2019 hasta mediados de enero de 2020 y se combinaron para obtener las mejores capturas. Los equipos utilizaron técnicas de segmentación de imágenes para ayudar en este proceso;
3. Las imágenes se recortaron en trozos de 100x100 píxeles con ejemplos positivos y negativos de granjas solares para entrenar el algoritmo, que utilizó una red neuronal convolucional (CNN) para distinguir las granjas solares de cualquier otro patrón negativo, como áreas agrícolas y montañosas, ciudades y suelo desnudo. De esta forma, el algoritmo podría distinguir patrones como bordes y combinaciones de colores con mayor precisión;
4. Finalmente, los equipos puntuaron su modelo en imágenes de todo el país, en particiones de 100x100 píxeles. Se procesaron un total de siete millones de imágenes.

Uno de los “mayores desafíos en el mapeo de parques solares es la similitud con los mismos” (Federico Bayle, director ejecutivo de Dymaxion Labs, febrero de 2020). Para solucionar esto, la CNN aplicada, que comprendía sistemas informáticos vagamente inspirados en las redes neuronales biológicas del

cerebro humano, consideró los colores en los canales RGB (rojo, verde y azul) de las imágenes; en este caso se consideró el canal infrarrojo ya que la distinción de los campos de cultivo era mejor.

1. Para una mayor autonomía, el robot debe usar una buena batería y debe ser recargable.
2. El diseño debe ser atractivo, para provocar interés en el niño.
3. La programación debe ser por bloques, como por ejemplo: Scratch.
4. El instructivo de armado debe ser ilustrativo, junto con la configuración del cableado.

En total, se utilizaron 10.999 imágenes para entrenar el modelo; El 70

- La precisión del algoritmo fue cercana al 94
- Se procesaron siete millones de imágenes;
- Se analizaron 2.780.400 kilómetros cuadrados;
- Se realizaron treinta horas de capacitación modelo.

Por lo tanto, estos resultados demuestran un conocimiento técnico que es particularmente importante en la industria de la IA. La falta de personal calificado necesario para programar los complejos sistemas y recursos técnicos fue un obstáculo importante. Este obstáculo surgió de los problemas socioeducativos observados al implementar soluciones de IA en cualquier industria que requiera importantes inversiones de capital para la capacitación y el desarrollo del personal. Al final, el proyecto

4. Conclusiones.

En este estudio, describimos el proceso de uso de CV, un subconjunto de ML e IA, en los informes de noticias, podemos notar cuatro grandes desafíos a la hora de implementar la IA, especialmente en proyectos que involucran imágenes satelitales y CV en las redacciones. En primer lugar, las imágenes de alta resolución son muy caras para cualquier medio. En el proceso algunos aprendimos una pincelada (introducción) a programas 3D, programación básica en Arduino, circuitos eléctricos, corte con máquina laser, a medir con pie de metro, etc. Para innovar en la industria de las noticias, los líderes de las redacciones deben gestionar procesos de innovación para activar y mejorar las capacidades creativas de sus equipos (Tidd and Bessant, 2009), como lo hizo LN Data al experimentar con CV para mejorar su capacidad de desarrollar habilidades basadas en datos y tomar sus informes al siguiente nivel. Esto amplía la teoría de la IA en la producción de noticias al proporcionar una nueva perspectiva sobre el estudio de la aplicabilidad y limitaciones del CV, principalmente en América Latina. Este tipo de

trabajo tiene limitaciones inherentes. Para concluir, es crucial comprender cómo se puede aplicar el CV a otras áreas de la sala de redacción y si será necesario establecer equipos dedicados al desarrollo de IA y ML en las organizaciones de medios de noticias. Para aprovechar las utilidades de la IA, las empresas de medios deben afrontar varios desafíos. Así, este estudio tuvo como objetivo proporcionar información sobre el desarrollo de las competencias necesarias para la integración del CV en proyectos implementados en las redacciones. Hasta donde sabemos, ningún estudio previo ha investigado el uso del CV en la producción de noticias.

5. Referencias.

Appelgren, E. y Nygren, G. (2014). Periodismo de datos en Suecia. *Digital Periodismo*, 2(3), 394–405. <https://doi.org/10.1080/21670811.2014.884344>

Appelgren, E., and Salaverría, R. (2018). La promesa de la transparencia Ency Culture: Un estudio comparativo del acceso a los datos públicos en las redacciones españolas y suecas. *Práctica del periodismo*, 12(8), 986–996. <https://doi.org/10.1080/17512786.2018.1511823>

Aronson, JD (2018). Visión por computadora y aprendizaje automático para humanos Análisis de vídeos sobre derechos: estudios de casos, posibilidades, preocupaciones y limitaciones. *Derecho e investigación social*, 43(4), 1188-1209. <https://doi.org/10.1111/lsi.12353>

Berliner, D. (2014). Los orígenes políticos de la transparencia. *Revista de política*, 76(2), 479–491. <https://doi.org/10.1017/S0022381613001412>

Biswal, SK y Gouda, NK (2020). Inteligencia artificial en el periodismo: ¿Una bendición o una pesadilla? *Optimización en Aprendizaje Automático y Aplicaciones* (págs. 155-167). Saltador. <https://doi.org/10.1007/978-981-15-0994-010>

Bonfanti, A. y Bordignon, M. (2017). 'Mariscos de esclavos': El Pulit- Premio zer a la luz de los Principios Rectores de las Naciones Unidas sobre las Empresas y los Derechos Humanos. *Política global*, 8(4), 498–504. <https://doi.org/10.1111/1758-5899.12495>