

La incorporación de la Ciencia de Datos como estrategia transversal para refozar la enseñanza de la programación en el aula

Sofía Martín¹, Claudia Banchoff¹, Paula Venosa¹, Liliana Hurtado²

¹LINTI - Laboratorio de Investigación en Nuevas Tecnologías Informáticas.
Facultad de Informática. Universidad Nacional de La Plata
Calle 50 esq. 120, 2do Piso. Tel: +54 221 4223528
{smartin, cbanchoff, pvenosa}@info.unlp.edu.ar}

²Facultad de Ciencias Económicas, Jurídicas y Sociales. Universidad Nacional de Salta
lilianahur@hotmail.com

RESUMEN

La incorporación de contenidos de la Ciencia de la Computación en la educación de nivel obligatorio ha avanzado en los últimos años a través de diferentes experiencias y propuestas. Los conceptos más abordados de esta ciencia, se centran mayormente en el pensamiento computacional y la programación, a través de propuestas relacionadas a contenidos de la caja curricular o bien contenidos específicos del área de Informática.

La Ciencia de Datos, parte de la Ciencia de la Computación, ha incrementado su importancia en áreas tales como la industria, ciencia y trabajo, debido a que se utiliza como base al momento de analizar situaciones y definir medidas a tomar. Es por esto que la comprensión al momento de trabajar con un conjunto de datos cobra importancia para luego realizar un correcto análisis.

Luego de haber analizado algunas de las experiencias implementadas en distintos niveles de enseñanza secundaria de nuestro región se pudo relevar que la Ciencia de Datos aún no se ha incluido como contenido de las propuestas pedagógicas en el sistema educativo de nivel secundario argentino. Por lo tanto, esta línea de investigación se centra en el análisis de experiencias pedagógicas sobre la enseñanza de la Ciencia de Datos en ámbitos educativos de nivel secundario, así como las herramientas y metodologías utilizadas, con el objetivo de generar propuestas transversales de enseñanza de esta área en las escuelas secundarias de la Provincia de Buenos Aires. Se trabaja

especialmente con herramientas basadas en el lenguaje de programación Python.

Palabras clave: Ciencias de la Computación, Ciencia de Datos, nivel secundario, Python, Jupyter notebook

CONTEXTO

En el Laboratorio de Investigación en Nuevas Tecnologías Informáticas (LINTI) se viene trabajando en diferentes proyectos relacionados a la incorporación de la Ciencia de la Computación en la escuela desde el año 2006. Algunos de los proyectos de extensión como “Programando con Robots y Software Libre” [1][2] y los proyectos que involucran contenidos de la Ciencia de la Computación nos han permitido estar en continuo contacto con escuelas para el armado y realización de distintas propuestas pedagógicas.

Las propuestas llevadas a cabo son variadas en cuanto a contenidos abordados. Generalmente se realizan en función de la posibilidad de acceso a los recursos con que cuenta cada institución y la heterogeneidad de las formaciones docentes. En particular, en el nivel secundario, su adaptación e incorporación depende de la modalidad de orientación del establecimiento.

La línea de investigación en este artículo se focaliza en un contenido que aún no es ampliamente abordado en las escuelas secundarias, especialmente en nuestra región,

como la Ciencia de Datos, que cada día cobra más relevancia para su enseñanza.

Esta línea de investigación se encuentra enmarcada en el proyecto: “De la Sociedad del Conocimiento a la Sociedad 5.0: un abordaje tecnológico y ético en nuestra región”, del Programa Nacional de Incentivos a docentes-investigadores, que se desarrolla en el LINTI.

1. INTRODUCCIÓN

La incorporación de los contenidos de la Ciencia de la Computación en las instituciones educativas de nivel secundario ha sido muy diferente en cada una, debido a los recursos con que cuenta cada establecimiento. El avance de este proceso se ve muy relacionado con los planes gubernamentales, nacionales y provinciales, los cuales llevan a cabo la entrega de diferentes tipos de recursos tecnológicos y propuestas pedagógicas para su utilización.

En el nivel secundario, estos contenidos que han sido incorporados no incluyen conceptos relacionados a la Ciencia de Datos, a diferencia de otros niveles educativos como el universitario, donde se introducen cada vez más carreras, especializaciones y cursos de perfeccionamiento en esta área.

El crecimiento de los datos en los últimos años, debido al amplio uso de dispositivos digitales e Internet, ha puesto en evidencia la importancia del análisis y los diferentes usos que pueden realizarse de los mismos. Los objetivos de uso de los datos es muy amplio, dado que los datos surgidos de las interacciones individuales de los dispositivos digitales generan posibilidades de adaptación y orientación de contenido, mientras que en el ámbito de la industria, ciencia y educación los tipos de análisis se enfocan de diferente manera[3]. Un estudio realizado desde

el sitio Statista¹ estima que durante el año 2020 se han generado 64.2 zettabytes de datos y este número sigue en crecimiento en forma exponencial cada año, se estima un crecimiento hasta los 181 zettabytes para el año 2025.

Este incremento de datos potenció la búsqueda de trabajos orientados al análisis de los mismos, y pone en juego habilidades que se pueden trabajar desde los niveles educativos previos a la universidad.

Algunos ámbitos de uso que se pueden nombrar son el análisis de cambios climáticos a lo largo de los últimos años, datos de enfermedades contrarrestados con los síntomas, en las administraciones de los gobiernos para generar sus políticas públicas a implementar y como forma de transparencia al compartir datos no sensibles en el manejo presupuestario, entre otras. El acceso a los datos generados por las políticas gubernamentales también brindan la posibilidad de abordar temas relacionados a los Objetivos de Desarrollo Sostenibles (ODS) propuestos por la ONU[4]. Desde un punto de vista similar podemos encontrar organizaciones como el sitio World Bank², que publican datos para su análisis o bien como fuente para generar desarrollos informáticos privados que permiten brindar distintos servicios como ser aplicaciones de tránsito, de clima, entre otras. Podemos encontrar otros proyectos en que se utilizan y publican datos abiertos como es el caso de DataGénero³, que visualizan situaciones problemáticas sobre perspectiva de género y a través del análisis y una buena comunicación de los mismos permiten poner en discusión dichas situaciones.

Asimismo como los contenidos de la Ciencia de la Computación ponen en juego habilidades importantes como el pensamiento crítico y la resolución de problemas, el proceso que implica el análisis de datos, facilita trabajar estas

¹ Link del sitio Statista:

<https://www.statista.com/statistics/871513/worldwide-data-created/>. Último acceso 07-03-2022

² Sitio con datos abiertos :

<https://databank.worldbank.org/home>. Último acceso 22-09-2021.

³ Sitio de DataGénero <https://linktr.ee/datagenero>.

Último acceso 22-09-2021.

habilidades en la escuela secundaria, además de generar una motivación adicional por tratarse de problemas del mundo real que permiten tomar decisiones críticas y pensar en formas de agrupar y visualizar los datos[5].

La experiencia a lo largo de los años de trabajar en proyectos en conjunto con escuelas nos permitió visualizar el desafío que conlleva la incorporación de contenidos que no son propios de la caja curricular, para lo cual es importante tener en cuenta varios aspectos, tales como la adaptación del contenido al nivel educativo y la adecuación de la herramienta a trabajar. Una de las aplicaciones informáticas que permite trabajar con análisis de datos que abordaremos es Jupyter Notebook, la cual facilita el trabajo de forma simplificada e interactiva y es una herramienta ampliamente utilizada[6]. En una primera instancia de trabajo se propone el uso de herramientas de escritorio visuales más simples como PandasGUI⁴ y PandaStable⁵ que simplifican el manejo de las operaciones de análisis de datos, para luego incorporar en las actividades el lenguaje de programación Python a través del uso de Jupyter notebook y de librerías tales como pandas y matplotlib.

2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

Esta línea de investigación se centra en los siguientes temas:

- Análisis sobre el impacto de la incorporación de conceptos de la Ciencia de Datos en la enseñanza transversal en la escuela.
- Relevamiento de distintas experiencias de enseñanza de la Ciencia de Datos en escuelas en otros países, incluyendo metodologías y herramientas de licencia libre utilizadas. En este sentido, se

analizarán las posibles adecuaciones de estas propuestas teniendo en cuenta las diferencias tanto de la caja curricular propia de cada jurisdicción como del nivel de contenidos específicos de Ciencia de la Computación ya abordados anteriormente.

- Diseño, elaboración y evaluación de materiales didácticos para la enseñanza de la Ciencia de Datos en el nivel secundario.
- Diseño, ejecución y evaluación de actividades específicas con estudiantes y docentes de nivel secundario.

Es importante destacar que este equipo de investigación viene interviniendo en distintas actividades en las escuelas de la región de La Plata, Berisso, Ensenada y Magdalena. También ha integrado el equipo que trabajó en el diseño curricular de la carrera “Especialización docente en didáctica de las Ciencias de la Computación” de la provincia de Buenos Aires y su implementación, y en la elaboración del “Primer manual de didáctica de las Ciencias de la Computación”⁶ ha generado experiencia en cuanto a poder intervenir en formación docente de nuestra provincia [8].

3. RESULTADOS Y OBJETIVOS

El objetivo principal de esta línea de trabajo es evaluar y adecuar herramientas de software que permitan elaborar propuestas de actividades que posibiliten introducir aspectos de la Ciencia de Datos, conforme a la caja curricular de las escuelas de nivel secundario pertenecientes a la Provincia de Buenos Aires, utilizando opciones de licencia libre.

Para alcanzar este objetivo, se proponen los siguientes objetivos específicos:

⁴ Disponible en :

<https://github.com/adameroose/pandasgui>

⁵ Documentación de PandaStable:

<https://pandastable.readthedocs.io/en/latest/description.html>

⁶ Este manual está disponible en

https://program.ar/descargas/cc_para_el_aula-2do_ciclo_secundaria.pdf . Último acceso: marzo de 2022

- Releva las diferentes experiencias pedagógicas desarrolladas en la región para la enseñanza de contenidos de la Ciencia de la Computación, área que integra la Ciencia de Datos, de manera tal de identificar los diferentes planes y programas que fueron acompañando la realización de las propuestas pedagógico-didácticas en la región.
- Releva las metodologías utilizadas y las herramientas propuestas para la enseñanza de contenidos específicos de Ciencias de la Computación y Ciencia de Datos, en las escuelas de nivel secundario.
- Investigar las diferentes opciones de las herramientas disponibles de licencia libre para el análisis y visualización de datos e identificar sus características y comparar uso, funcionalidad y posibles adecuaciones para escuelas de la región.
- Desarrollar propuestas pedagógicas para la enseñanza de la Ciencia de Datos en la escuela secundaria.
- Diseñar, implementar y evaluar las intervenciones con docentes y estudiantes de nivel secundario.
- Habilitar un espacio de aprendizaje formativo destinado a los docentes de los establecimientos del nivel secundario, con el fin de que utilicen y transfieran los resultados de la investigación en sus propuestas pedagógicas.
- Generar un sitio web público de referencia con ejemplos de análisis de datos adaptados a la escuela secundaria y brindar tutoriales de uso de las diferentes herramientas.

Durante el año 2022 se comenzó a trabajar con la escuela EET 9 de la ciudad de La Plata. Estas actividades permitieron definir una serie de acciones y ajustes a la propuesta inicial. Se analizaron herramientas visuales alternativas a Jupyter Notebook de manera de poder introducir los conceptos básicos de una manera más

simple para luego realizar actividades con el lenguaje de programación Python y librerías como pandas y matplotlib⁷.

La fig. 1 muestra capturas de PandaStable y PandasGUI. Estas herramientas fueron probadas en el sistema operativo Huayra Linux, distribución provista por las netbooks disponibles en la mayoría de los establecimientos educativos de la provincia de Buenos Aires.

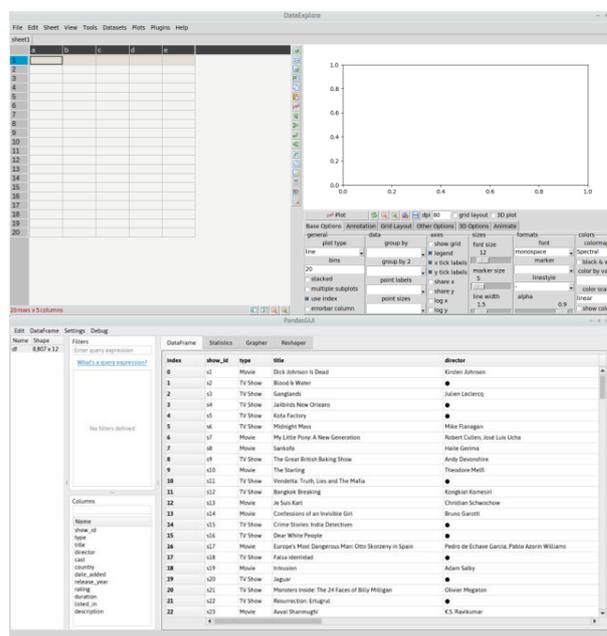


Figura 1: imagen de PandaStable y PandasGUI. Una primera propuesta se presentó como demo educativa en TEyET 2022[9] la cual tuvo una gran aceptación por parte de los docentes participantes.

4. FORMACIÓN DE RECURSOS HUMANOS

El equipo de trabajo de la línea de I+D+i presentada en este artículo está formado por docentes investigadores categorizados del LINTI. Si bien el enfoque hacia la Ciencia de Datos es una línea que recién inicia, tiene como antecedentes años de investigación en trabajo

⁷ <https://matplotlib.org/stable/index.html>

con escuelas presentados en diversos ámbitos [10][11].

Esta línea de investigación dió lugar al desarrollo de una tesis de postgrado y actividades de cátedras relacionadas con esta temática, como así también el proyecto de extensión denominado “Ciencia de Datos en la escuela” que permitió obtener evidencias con base empírica para continuar indagando sobre las formas de incorporar esta disciplina en la educación media.

5. BIBLIOGRAFÍA

- [1] Claudia Queiruga, Claudia Banchoff Tzancoff, Paula Venosa, Sofia Martin, Vanessa Aybar Rosales, Soledad Gomez, Isabel Kimura. *EscuelasTIC: estrategias para trabajar el pensamiento*. XIII Edición del Workshop de Investigadores en Ciencias de la Computación. Chilecito : UNdeC, 2021. ISBN 978-987-24611-3-3.
- [2] Banchoff Tzancoff, C. M., Martin, E. S., Gómez, N. S., & López, F. E. M. (2019). Experiencias en robótica educativa: Diez años trabajando con escuelas argentinas. XIV Congreso Nacional de Tecnología en Educación y Educación en Tecnología (TE&ET 2019) (Universidad Nacional de San Luis, 1 y 2 de julio de 2019). ISBN 978-987-733-196-7, pág. 84-93. <http://sedici.unlp.edu.ar/handle/10915/90729>.
- [3] Rao, A. R., Desai, Y., & Mishra, K. (2019). Data science education through education data: An end-to-end perspective. *2019 IEEE Integrated STEM Education Conference (ISEC)*, 300-307. <https://doi.org/10.1109/ISECon.2019.8881970>
- [4] Departamento de Asuntos Económicos y Sociales (DESA). Naciones Unidas. Informe de los Objetivos de Desarrollo Sostenible 2020. e-ISBN: 978-92-1-004963-4 ISSN: 2521-6899 e-ISSN: 2521-6902. Disponible en <https://unstats.un.org/sdgs/report/2020/>
- [5] De Veaux R, Agarwal M, Averett M, et al. Curriculum Guidelines for Undergraduate Programs in Data Science. *Annu Rev Stat Appl*. 2017;4(1):15-30. doi:10.1146/annurev-statistics-060116-053930
- [6] Randles BM, Paschetto IV, Golshan MS, Borgman CL. Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study. In: *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*; 2017:1-2. doi:10.1109/JCDL.2017.7991618
- [7] Díaz Barriga F. (2003). *Cognición situada y estrategias para el aprendizaje significativo*. Revista. Electrónica de Investigación Educativa, 5(2). Disponible en: <https://bit.ly/1JSJZpG>. Último acceso marzo de 2021.
- [8] Martin, S., Banchoff, C., Venosa, P., & Hurtado, L. (2022). *Ciencias de datos en escuelas secundarias*. XVII Congreso de Tecnología en Educación & Educación en Tecnología - TE&ET 2022 (Entre Ríos, 15 y 16 de junio de 2022). <http://sedici.unlp.edu.ar/handle/10915/139959>
- [9] Queiruga C., Banchoff Tzancoff C., Gómez S., Venosa P. (2020). *Enseñar a Enseñar Ciencias de la Computación. Una experiencia sobre políticas educativas y contenidos de Ciencias de la Computación*. En Actas II Jornadas Argentinas de Didáctica de la Programación / Alejandro Iglesias... [et al.]; editado por Araceli Acosta... [et al.]- 1a ed.- Córdoba: Universidad Nacional de Córdoba. Facultad de Filosofía y Humanidades, 2020. ISBN 978-950-33-1600-9
- [10] Queiruga, C. A., Banchoff Tzancoff, C. M., Venosa, P., Martín, S. S., Aybar Rosales, V. del C., & Soledad Gomez, I. K. (2021). *EscuelasTIC: Estrategias para trabajar el pensamiento computacional en la escuela argentina*. XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021, Chilecito, La Rioja). ISBN 978-987-24611-3-3. <http://sedici.unlp.edu.ar/handle/10915/120472>
- [11] Queiruga, C. A., Banchoff Tzancoff, C. M., Venosa, P., Martin, E. S., Aybar Rosales, V. del C., Gómez, N. S., & Kimura, I. *EscuelasTIC. El pensamiento computacional en la escuela*. XXI Workshop de Investigadores en Ciencias de la Computación (WICC 2019, Universidad Nacional de San Juan).