

Ciencias de datos en escuelas secundarias

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

Ámbito de aplicación

Las actividades a mostrar están pensadas para ser aplicadas en el ámbito de la escuela secundaria.

Esta demo presenta un trabajo en el marco de la línea de investigación “Espacio de aprendizaje de la Ciencia de Datos para establecimientos educativos de la Provincia de Buenos Aires”, el cual se trabaja también en el proyecto “Ciencia de Datos en la escuela subsidiado” subsidiado por la Secretaría de Políticas Universitarias en la convocatoria de 2021.

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 de esta ciencia, que se han ido implementando, 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 se ha vuelto un tema de interés en diferentes áreas de estudio, dado que el correcto análisis de datos se torna un aporte importante al momento de identificar situaciones actuales, definir las acciones a tomar, encontrar posibles problemas según el ámbito de estudio.

Se han analizado algunas de las experiencias implementadas en distintos contextos y se pudo relevar que la Ciencia de Datos aún no se ha incluido como contenido de las propuestas educativas.

Objetivos

El objetivo general de la propuesta es incorporar contenidos de la Ciencia de la Computación, en particular Ciencia de Datos en el nivel secundario.

En esta demo se plantean una serie de actividades cuyo objetivos son:

- Mostrar herramientas disponibles de licencia libre, identificar sus características y comparar uso, funcionalidad
- Presentar casos de aplicación orientados a diferentes áreas de la currícula de nivel secundario.
- Concientizar en el uso de datos abiertos tanto para su uso como para su generación.
- Detallar el proceso de análisis recursivo con el fin de obtener, curar, gestionar y procesar datos que permitan definir preguntas y comunicar los resultados.

En este sentido, todos los recursos utilizados, como las propuestas realizadas son de libre distribución y se encuentran disponibles para su utilización.

Descripción

El desafío de incorporar contenidos que no son propios de la caja curricular del Nivel Secundario incluye varios aspectos a tener en cuenta, tales como adaptar el contenido al nivel

educativo y la adecuación de la herramienta a trabajar. Una de las aplicaciones informáticas más populares, que permite trabajar con análisis de datos es Jupyter Notebook, la cual mostraremos en esta demo. Esta herramienta facilita el trabajo de forma simplificada e interactiva . [1]

Los establecimientos educativos han ido incorporando contenidos de Ciencia de la Computación en forma muy dispar y su avance está relacionado con la implementación de planes gubernamentales, nacionales y provinciales, que incluyen la entrega de diferentes tipos de recursos tecnológicos acompañados de propuestas pedagógicas. En particular la Ciencia de Datos no es aún tenida en cuenta como contenido a trabajar dentro de la escuela, si bien cada día se pone en evidencia su creciente importancia a través de generación de carreras, especializaciones universitarias, es escasa su implementación en nivel secundario.

La Ciencia de Datos ha ido tomando relevancia en los últimos años debido a la gran cantidad de datos que se generan a través del uso masivo de los dispositivos digitales y sitios en Internet. Su utilización ha crecido no sólo a nivel individual en las redes sociales sino también en ámbitos de trabajo, científico, entre otros[2]. Un estudio publicado en el sitio Mirabaud¹ estima que en el período entre 2021-2024 la generación de datos se incrementará en un 26% anual.

Su crecimiento ha generado el interés de análisis para potenciar actividades que mejoren aspectos de producción en la industria, establecer patrones de enfermedades y poder realizar diagnósticos y estudios más eficientes en la ciencia, en función de datos recolectados.

Otro ámbito de implementación en el área gubernamental que no solo utilizan la Ciencia de

Datos para una análisis interno, sino que también al publicar los datos no sensibles como medida de transparencia, generan la posibilidad de que las organizaciones, como el Banco Mundial², puedan utilizarlos y generar desarrollos informáticos privados para brindar servicios. En algunos casos se utilizan por organizaciones orientadas a problemáticas específicas que visibilizan situaciones problemáticas, que cobran mayor fuerza a través de la comunicación significativa con datos. Un ejemplo de esto último es el sitio de DataGénero³ que compila información relacionada a la desigualdad de género en nuestro país. 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[3]. La introducción de este tipo de análisis en la escuela media permite generar propuestas para fomentar el pensamiento crítico de los estudiantes y entregar herramientas que permitan una mejor toma de decisiones a los equipos docentes para el seguimiento de sus estudiantes[4].

En la demo se mostrarán algunos ejemplos en los cuales se trabaja con datos abiertos y posibles formas de visualización de los mismos. La Figura 1, muestra un típico gráfico de torta que toma información de un dataset que proviene de la base de datos nacional de nutrientes del USDA (Departamento de Agricultura de Estados Unidos)⁴. Estos gráficos y otros similares, son creados con la herramienta Matplotlib⁵, la cual es muy utilizada en este ámbito.

¹ THE EXPLOSION OF DATA:
https://www.mirabaud-am.com/fileadmin/user_upload/equities/mirabaud-am-the-explosion-of-data.pdf. Último acceso 10 de Abril 2022

² Sitio con datos abiertos :
<https://databank.worldbank.org/home>. Último acceso 22-09-2021.

³ <https://linktr.ee/datagenero>. Último acceso 22-09-2021.

⁴ <https://data.world/exercises/principal-components-exercise-1>

⁵ Sitio oficial: <https://matplotlib.org/>

Los 5 grupos de comida con mayor cantidad de calorías

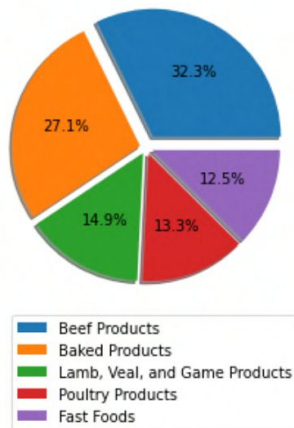


Figura 1. Gráfico de torta generada por Matplotlib

Otra posible forma de visualizar datos es a través de mapas de color, donde a través de una paleta de colores se presentan proporciones de los valores correspondientes al set de datos como podemos ver en la Figura 2. Los datos en la figura representan datos geográficos de la autoría del IGN (Instituto Geográfico Nacional de la República Argentina). Estos mapas se generan a través de la librería GeoPandas⁶.

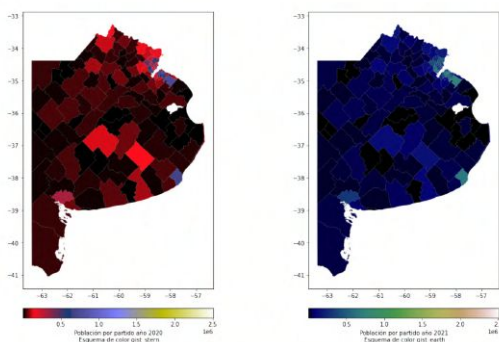


Figura 2. Mapa de la provincia de Buenos Aires generado por geopandas.

Como un último ejemplo, la Figura 3 muestra una nube de palabras generadas con la herramienta wordcloud⁷ a partir de datos obtenidos de la plataforma Spotify sobre música latina.



Figura 3. Nube de palabras generada por wordcloud.

Los materiales incluidos en esta demo se encontrarán disponibles para su descarga. Para replicar esta demo es necesario tener instalado la herramienta Jupyter Notebook en forma local o bien, puede utilizarse la herramienta Google Colab.

Referencias

- [1] Randles BM, Pasquetto 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
- [2] A. Ravishankar Rao, Yashvi Desai, and Kavita Mishra. Data science education through education data: An end-to-end perspective. In *2019 IEEE Integrated STEM Education Conference (ISEC)*, pág. 300–307. IEEE.
- [3] 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/>
- [4] 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 Sitio oficial: <https://geopandas.org/en/stable/>

⁷ Repositorio: https://github.com/amueller/word_cloud