
Sobre la propuesta de un programa de matemáticas aplicadas en ciencia de datos

About the proposal of a bachelor of applied mathematics in data science degree

Juan Triana^a
cbasicas@uniagustiniana.edu.co

Resumen

Este artículo presenta los resultados del estudio realizado por el Departamento de Ciencias Básicas de la Universitaria Agustiniiana, con el fin de determinar la pertinencia de un programa de pregrado de matemáticas aplicadas en ciencia de datos. Se realizó una encuesta abierta a la comunidad para determinar el interés en diversos programas de ciencias básicas, en modalidad virtual y presencial, y para conocer qué tan relevante consideran algunos aspectos formativos entre ellos el uso de herramientas de software; por otra parte, se hizo un análisis de ofertas laborales relacionadas con análisis y ciencia de datos, para conocer los perfiles y herramientas de software solicitados en el mercado laboral.

Palabras clave: Matemáticas aplicadas, ciencia de datos, educación superior, encuesta..

Abstract

We present the results of a quantitative study by the Department of Basic Science of the Universitaria Agustiniiana, to determine the relevance of a bachelor of applied mathematics in data science degree. An open survey was performed to assess interest in several Bachelor of Science degrees, in virtual and face-to-face modalities, and to know how relevant they consider some aspects such as the use of software tools in professional practice. On the other hand, job offers related to data analysis and data science were analyzed, to know the profiles and software tools requested in the labor market.

Keywords: Applied mathematics, data science, higher education, survey..

^aUniversidad Agustiniiana

1. Introducción

La comprensión del mundo real ha sido fundamental para el desarrollo de las matemáticas, pues diversos conocimientos matemáticos surgieron como solución a las necesidades básicas de la humanidad mediante procesos de abstracción ya que, como indica el matemático y divulgador científico Ian Stewart, el progreso de la civilización humana y el progreso de las matemáticas han ido de la mano (Stewart 2008). El avance en los diferentes campos de conocimiento y las necesidades del sector productivo traen nuevos desafíos que imponen una necesidad de reorganizar los saberes; en 1962 Jhon W. Tukey (1915-2000), reconocido por ser el precursor de los box plot (Nuzzo 2016), por el lema de Tukey de teoría de conjuntos por sus trabajos en series de tiempo y por la transformada rápida de Fourier, entre otros aportes (Brillinger 2002), publicó el artículo *The Future of Data Analysis* en el cual describe diversos cambios propiciados por los desarrollos en la estadística matemática como también por el surgimiento del computador, además señala la necesidad de adaptar la forma en que se abordaba la estadística en la época, indicando la existencia de un área de estudio cuyo objetivo sería el aprendizaje a partir de los datos (Tukey 1962).

Por otra parte, en 1974 Peter Naur (1928 - 2016) propone ciencia de datos como nombre alternativo para las ciencias de la computación, enfatizando en la importancia de las herramientas que se empleen para procesar y analizar los datos (Naur 1974). En 1977 surge *The International Association for Statistical Computing* (IASC), con el fin de vincular la metodología estadística tradicional y la tecnología informática, fomentando el intercambio de conocimiento mediante eventos internacionales que integran a estadísticos, profesionales de la computación, organizaciones, instituciones, gobiernos y el público en general.

El desarrollo tecnológico permitió la implementación de algoritmos para la realización de procesos complejos, además de la posibilidad de generar, procesar y analizar grandes volúmenes de datos, lo que además ha llevado a una constante evolución en las herramientas y métodos. Un ejemplo de esta evolución lo tenemos en el Big Data, iniciando en 2001 con una caracterización de Doug Laney a través de las 3 V: volumen, velocidad y variedad que hacen referencia al tamaño de los conjuntos de datos, la velocidad de generación y procesamiento de datos, y a la diversidad en los tipos y complejidad de los datos, respectivamente. Posteriormente recibe una caracterización con 5V, agregando veracidad y valor, en donde veracidad hace especial énfasis en la confiabilidad de los datos, y valor hace referencia a la capacidad de transformar los datos en información (Husamaldin & Saeed 2020). Actualmente, se suele considerar un modelo de 7 V en el cual a las características mencionadas se adicionan visualización, refiriéndose a la forma en la que se presentan los datos, y viabilidad en la cual se indaga sobre la capacidad que tenga la compañía para manejar y gestionar los datos (Niculescu 2020); sin embargo, dependiendo de sus necesidades, hay autores que consideran incluso 10 V en sus trabajos (Bhat & Huang 2021).

La matemática aplicada ejerce como una conexión entre lo teórico y lo práctico con

lo cual, mediante métodos y objetos matemáticos, es posible abordar y representar problemas aplicados provenientes de otras disciplinas. Se debe tener presente que la ciencia de datos es un área multidisciplinaria que combina conocimientos en modelación matemática y estadística, dominio de diversas herramientas de software y conocimiento del contexto correspondiente a los datos para asegurar una apropiada toma de decisiones, como lo muestra la siguiente Figura tomada de (Sánchez et al. 2022).



Figura 1: Diagrama de Conway de la ciencia de datos.

Incluso si se considera a la ciencia de datos como un área independiente, es innegable que esta debe tener como base científica a las matemáticas (Nasution et al. 2020). Teniendo en cuenta la Figura 1, un programa que parta de la base de las matemáticas aplicadas puede aportar a la generación de conocimiento en la ciencia de datos partir de la construcción de modelos, simulación, uso de métodos de machine learning e investigación tradicional, e incluso técnicas específicas de la descomposición de matrices (Cabrera & Triana 2021) y de otras ramas de las matemáticas (Nield 2022), como se evidencia en la siguiente Figura tomada de (Deisenroth et al. 2020).

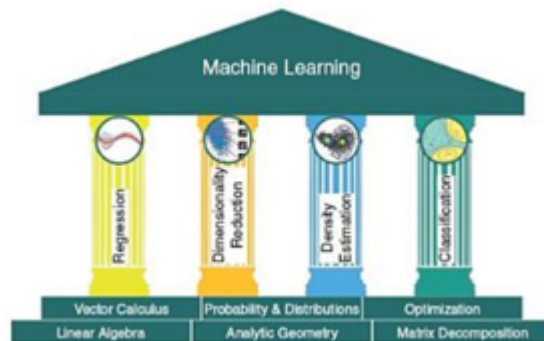


Figura 2: Fundamentos matemáticos de los métodos de machine learning.

La pertinencia de un programa de matemáticas aplicadas en ciencia de datos, a nivel nacional, es dada por la necesidad de diversos sectores de contar con profesionales con capacidad para el análisis de datos y la construcción de modelos;

como se observa en el estudio (Alianza TIC 2020), el cargo analista de datos se encuentra en el Top 5 de los cargos críticos y científico datos es el cargo que más aumentará su demanda en el futuro, en la siguiente Figura se muestran algunos de los resultados presentados en (Alianza TIC 2020)

Top 5 cargos críticos en la empresa	Top 5 cargos nuevos	Top 3 cargos que aumentarán su demanda en el futuro
Desarrollador	Profesional en Minería de Datos	Científico de Datos
Consultor	Analista de Blockchain	Administrador de Servicios en la Nube
Analista de Datos	Entrenador de sistemas expertos	Arquitecto TI
Arquitectos	Especialistas en seguridad	
Analista de Soporte	Administrador de dispositivos	

Figura 3: Top de cargos relacionados con las TIC.

Sin embargo, los perfiles de ciencia de datos y análisis de datos requieren de mayor rigor analítico dado que entre sus funciones se encuentra el extraer información de los datos, el uso de técnicas de modelado para la solución de problemas, la creación de modelos analíticos, entre otros (IBM 2017); lo que evidencia la pertinencia de un enfoque con base en la fundamentación conceptual que ofrecen las matemáticas pero con una orientación a la aplicación del conocimiento, que es lo que se busca en un programa de matemáticas aplicadas en ciencia de datos.

DSA Framework Category	Functional Role	Sample Occupations
 Data Scientists & Advanced Analytics	Create sophisticated analytical models used to build new datasets and derive new insights from data	Data Scientist Economist
 Data Analysts	Leverage data analysis and modeling techniques to solve problems and glean insight across functional domains	Data Analysts Business Intelligence Analyst
 Data Systems Developers	Design, build and maintain and organization's data and analytical infrastructure	Systems Analyst Database Administrator
 Analytics Managers	Oversee analytical operations and communicate insights to executives	Chief Analytics Officer Marketing Analytics Manager
 Functional Analysts	Utilize data and analytical models to inform specific functions and business decisions	Business Analyst Financial Analyst
 Data-Driven Decision Makers	Leverage data to inform strategic and operational decisions	IT Project Manager Marketing Manager

Figura 4: Tomado del estudio *The Quant Crunch* de IBM.

Un análisis bibliométrico con base en la búsqueda de la palabra data, realizado en (Mendoza et al. 2021), encontró que la mayoría de las palabras clave y temáticas relevantes se refieren a los métodos de la modelación de datos con algoritmos y a la gestión de tecnología para la administración de grandes bases de datos, lo que muestra la necesidad de hacer explícita la distinción entre los perfiles de ciencia de datos y de ingeniería de datos, tal como lo hace (IBM 2017) a través de las funciones de cada perfil.

En este trabajo se busca determinar la percepción sobre el programa académico de pregrado Matemáticas Aplicadas en Ciencia de Datos, y la viabilidad de este para ofertarse ya sea en modalidad presencial o en modalidad virtual en la Universitaria Agustiniiana. No sobra aclarar que, a nivel nacional, la Corporación Universitaria Autónoma del Cauca cuenta con un programa con dicha denominación mientras que a nivel internacional existen diversos programas con denominación similar.

2. Materiales y Métodos

Se realizó una encuesta voluntaria y anónima a través de un formulario electrónico de Google, la encuesta fue diseñada con preguntas cerradas en las que se indaga respecto a las expectativas educativas con relación a la modalidad, para lo cual se consideran las modalidades virtual y presencial; además, se consulta acerca de la preferencia sobre algunos posibles programas de ciencias básicas y se realizan algunas preguntas para determinar si consideran relevantes algunos aspectos formativos, como la construcción de modelos, el uso de herramientas de software y el análisis de datos, para la construcción de soluciones de problemáticas en sus lugares de residencia. En la encuesta participaron 536 personas de los cuales el 51.5% son mujeres y el 48.5% son hombres, el 94.4% de los encuestados habitan en Bogotá y Cundinamarca.

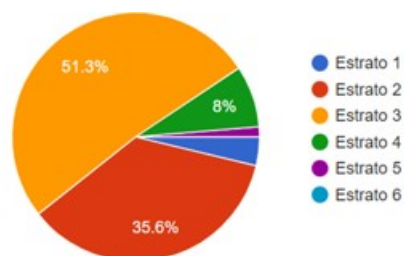


Figura 5: Estrato de los encuestados.

Como se observa, el 86.9% de los encuestados corresponden a los estratos 2 y 3. El perfil de los encuestados es variado, 41.4% son profesionales, 32.5% estudiantes universitarios, 11.2% estudiantes de colegio, 6.9% son estudiantes de programas técnicos o tecnológicos, 5.4% tiene formación técnica o tecnológica, el restante son graduados de bachillerato. El 39% de los encuestados, estudia o estudió en colegio privado, los demás fueron formados en colegios públicos; el 4.9% de los encuestados se encuentra en zona rural, los demás están ubicados en zonas urbanas.

Por otra parte, se realizó un análisis de las necesidades del mercado laboral mediante la recolección de 194 ofertas laborales relacionadas con análisis de datos, tomadas del portal empleo.com, en el periodo comprendido entre junio y diciembre de 2022. Los resultados permitieron determinar que los perfiles más solicitados eran estadísticos, ingenieros industriales, ingenieros de sistemas y matemáticos;

en muchos casos pese a que la oferta indicaba un perfil de analista de datos, sus funciones correspondían a un ingeniero de datos.



Figura 6: Profesiones más demandadas para cargos de análisis de datos.

3. Resultados

En la encuesta, además de recolectar información socioeconómica, se obtienen los siguientes resultados. En lo referente a los programas y a la modalidad académica, se emplean preguntas tipo Likert en donde 1 significa poco interés y 5 bastante interés.

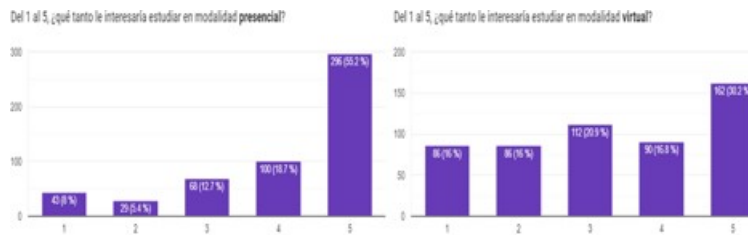


Figura 7: Consulta sobre la modalidad para realizar estudios de pregrado en áreas de Ciencias Básicas.

Se evidencia mayor interés por la modalidad presencial; no obstante, se puede apreciar que hay alto interés en contar con acceso a programas en modalidad virtual. En cuanto a los programas de Ciencias Básicas que pueden ofertarse en la Uniagustiniana, para el caso de programas relacionados con matemáticas se tiene en cuenta que, de acuerdo con el SNIES, para el año 2023 se reportan 69 programas de pregrado activos cuyo nombre incluye matemáticas, de los cuales el 35% tiene un enfoque en las matemáticas teóricas; por lo anterior, se consideró relevante consultar a la comunidad por su interés en un programa de matemáticas, o matemáticas puras, y por el programa de matemáticas aplicadas en ciencia de datos.

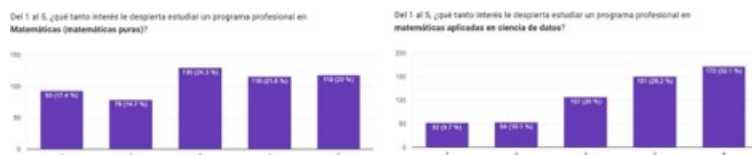


Figura 8: Respuestas sobre el interés en estudiar un programa profesional en matemáticas y en matemáticas aplicadas en ciencia de datos

Se observa un mayor interés en el programa de matemáticas aplicadas en ciencia de datos, pero no se debe ignorar que las matemáticas teóricas despiertan bastante curiosidad en la comunidad. Por otra parte, se consideraron los programas de biología, física y química como opciones obteniendo los siguientes resultados. Se evidencia un mayor interés en el programa de biología que en física y química, mostrando resultados similares a los obtenidos en el estudio realizado por la Universidad Surcolombiana en el sur del Huila, en donde se concluyó que los programas de matemática aplicada y biología eran la mejor opción como nuevos programas de ciencias básicas (Caicedo & Carrera 2017). En cuanto a lo referente a la relevancia de algunos aspectos formativos, se tienen los siguientes resultados.

Según (OIT 2017), el 36% de los jóvenes señaló que no aplican los conocimientos y destrezas adquiridas durante su formación en su trabajo; además, indica que estos porcentajes son mayores cuando se trata de programas académicos de Ciencia, Tecnología, Ingeniería y Matemática, que son programas necesarios para abordar la revolución tecnológica actual. Como muestra el estudio de empleo juvenil y emprendimiento en América Latina, desarrollado por el Banco Centroamericano de Integración Económica (BCIE) y la Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura (OEI) presentado en 2022, esta discrepancia entre las habilidades adquiridas y las necesidades del mercado laboral hacen que las empresas tengan dificultades para cubrir sus vacantes, como se evidencia en los documentos de Brechas de Capital Humano en los catálogos del Marco Nacional de Cualificaciones. Por lo anterior, resulta pertinente la propuesta del programa matemáticas aplicadas en ciencia de datos ya que se enmarca en la aplicación del conocimiento, en concordancia con las necesidades del mercado laboral en el contexto nacional, brindando espacios para fortalecer el análisis de datos para la toma de decisiones, la construcción de modelos, el uso de lenguajes de programación y herramientas de software, habilidades que tienen gran relevancia para los encuestados según la Figura 10.

Para determinar las herramientas de software y lenguajes de programación más adecuados para los espacios formativos para el aprendizaje que compondrían el programa de matemáticas aplicadas en ciencia de datos, se hizo un análisis de frecuencias con las ofertas laborales recolectadas con el cual, como se muestra en la Figura 11, se puede observar una mayor preferencia por el software libre.

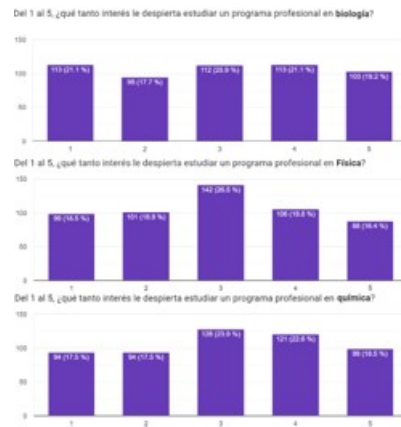


Figura 9: Respuestas sobre el interés en programas de biología, física y química.

4. Conclusiones

De los resultados obtenidos al consultar por la preferencia respecto a diversos programas de ciencias básicas, que se pueden apreciar en las Figuras 8 y 9, se concluye que el programa de matemáticas aplicadas en ciencia de datos despierta mayor interés, con lo cual puede considerarse como la opción más viable entre los programas consultados. De los resultados presentados en la Figura 11, se puede recomendar que a lo largo del proceso formativo se empleen como lenguajes de programación Python y R, ya que permiten al estudiante tener mayor adaptación al mercado laboral; no obstante, se sugiere realizar constantes análisis del mercado laboral para actualizar las herramientas y lenguajes de programación que serán empleados en los procesos formativos, esto dado el surgimiento de nuevo software que puede posicionarse en el mercado (Gao et al. 2020)

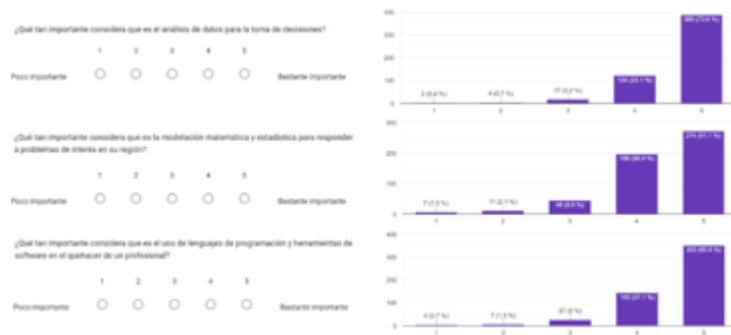


Figura 10: Respuestas sobre la importancia de determinadas aspectos formativos.



Figura 11: Herramientas de software más demandadas.

Teniendo en cuenta que, según el SNIES, en 2021 hubo 375086 estudiantes en modalidad virtual, representando un incremento del 49.2% respecto al año 2020, y dados los resultados de preferencia de modalidad presentados en la Figura 7, la modalidad virtual es una opción viable para el programa; además, esta modalidad permite ampliar la cobertura de la institución por las ventajas que ofrece para llevar el conocimiento a las diversas regiones del país, atendiendo a la necesidad de educación superior descrita en (Departamento Nacional de Planeación 2023).

El Departamento Nacional de Planeación (DNP), la Dirección de Ordenamiento y Desarrollo Territorial Sostenible (DODT) y su Subdirección de Planeación Territorial (SPT) han entregado al público el Sistema de Estadísticas Territoriales - TerriData, como respuesta a las recomendaciones de la Organización para la Cooperación y el Desarrollo Económico (OCDE), para promover la gestión de conocimiento de la operación pública a partir de datos estadísticos a nivel municipal, departamental y regional en Colombia. La consolidación de plataformas gubernamentales de datos libres permite que desde un programa de matemáticas

aplicadas en ciencia de datos se pueda aportar a la solución de problemas propios de las regiones, empleando datos reales para el abordaje de problemas en el aula, para la construcción de proyectos y la generación de productos de investigación, como se evidencia en (Mejía et al. 2021).

Recibido: 1 de octubre de 2023

Aceptado: 12 de diciembre de 2023

Referencias

Alianza TIC (2020), *Resultado del estudio de identificación de brechas de capital humano para el Sector TIC: con enfoque en la explotación de datos y prospectiva*, Ministerio del Trabajo, Ministerio de las TIC, Ministerio de Educación Nacional, Sena y Mesa sectorial.

Banco Centroamericano de Integración Económica (BCIE) and Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura (OEI)

- (2022), *Empleo juvenil y emprendimiento en América latina y el caribe*, Informe Completo.
- Bhat, S. & Huang, N. (2021), ‘Big data and ai revolution in precision agriculture: Survey and challenges’, *IEEE Access* **9**, 110209–110222.
- Brillinger, D. (2002), ‘John w. tukey’s work on time series and spectrum analysis’, *The annals of statistics* **30**(6), 1595–1618.
- Cabrera, S. & Triana, J. (2021), ‘Linear equality-constrained least-square problems by generalized QR factorization’, *Selecciones matemáticas* **8**(2), 437–443.
- Caicedo, L. & Carrera, N. (2017), ‘Estudio de factibilidad para la apertura del programa de matemática aplicada en la Universidad Surcolombiana sede Pitalito’, *Entornos* **30**(2), 73–77.
- Deisenroth, M., Faisal, A. & Ong, C. (2020), *Mathematics for machine learning*, Cambridge University Press, New York.
- Departamento Nacional de Planeación (2023), *Colombia potencia mundial de la vida: Bases del plan nacional de desarrollo 2022-2026.*, Bases del Plan Nacional de Desarrollo.
- Gao, K., Mei, G., Picciali, F., Cuomo, S., Tu, J. & Huo, Z. (2020), ‘Julia language in machine learning: Algorithms, applications, and open issues’, *Computer Science Review* **37**, 100254.
- Husamaldin, L. & Saeed, N. (2020), ‘Big Data analytics correlation taxonomy’, *Information* **11**(1), 1–12.
- IBM (2017), *The quant Crunch: how the demand for data science skills is disrupting the job market.*, Burning glass technologies.
- Mejía, M., Cardona, F. & Mejía, J. (2021), ‘El desarrollo endógeno en el municipio de San Sebastián de Buenavista Magdalena: Una alternativa ante los efectos del Covid-19’, *Cultura latinoamericana* **33**(1), 154–186.
- Mendoza, F., J., Q., Acevedo, O. & García, F. (2021), ‘Fundamentación teórica para la creación de un programa académico de ingeniería y ciencia de datos: una aplicación bibliométrica’, *Aibi revista de investigación, administración e ingeniería.* **9**(3), 49–58.
- Nasution, M., Sitompul, O. & Nababan, E. (2020), ‘Data Science’, *Journal of Physics: Conference Series* **1566**, 012034.
- Naur, P. (1974), *Concise Survey of Computer Methods*, Petrocelli Books.
- Niculescu, V. (2020), ‘On the impact of high performance Computing in Big Data analytics for medicine’, *Applied medical informatics* **42**(1), 9–18.
- Nield, T. (2022), *Essential math for data science*, O’Reilly.

- Nuzzo, R. (2016), 'The box plots alternative for visualizing quantitative data', *PM&R* **8**(3), 268-272.
- OIT (2017), *El futuro del trabajo que queremos. La voz de los jóvenes y diferentes miradas desde América Latina y el Caribe*, Oficina Regional para América Latina y el Caribe.
- Sánchez, P., Arcila, C. & Blanco, D. (2022), 'Conocimiento y percepción de la ciudadanía española sobre el big data y la inteligencia artificial', *Revista de comunicación y tecnologías emergentes* **20**(1), 1-25.
- Stewart, I. (2008), *Historia de las matemáticas en los últimos 10000 años*, Editorial crítica.
- Tukey, J. (1962), 'The future of data analysis', *The annals of mathematical statistics*. **33**(1), 1-67.