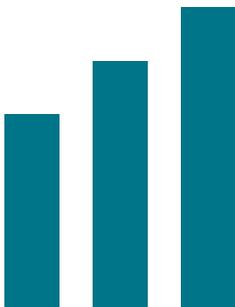


TENDENCIAS TECNOLÓGICAS EN EL ANÁLISIS DE DATOS: PATRONES A GRAN ESCALA O PISTAS ESPECÍFICAS DE GRAN VALOR, LA SIMBIOSIS ENTRE *BIG DATA* Y *SMALL DATA*

Artículo de investigación científica



Héctor Francisco Galeros Juárez

Ingeniero en Electrónica con maestría en Administración Industrial y de Empresas de Servicio. Universidad de San Carlos de Guatemala.

Correo electrónico: franciscogaleros@gmail.com

Fecha de recepción: 17/09/2018

Fecha de aceptación: 16/10/2018

Resumen

El *World Economic Forum*, en su noveno pilar denominado Adopción Tecnológica, ubica a Guatemala en el puesto 43 de 140 países. Dado que es un pilar importante para la competitividad empresarial, el siguiente artículo se concentra en la determinación de las tendencias tecnológicas y herramientas de análisis de datos utilizadas por estudiantes y exalumnos de la Universidad Rafael Landívar de Guatemala, debido a las tendencias actuales del *Big Data* y *Small Data* y su importancia en la inteligencia de negocios y la toma de decisiones a nivel empresarial.

Palabras clave: tendencia, tecnología, macro datos, pequeños datos, análisis de datos, patrones, pistas, herramientas tecnológicas.

Abstract

The World Economic Forum in its ninth pillar called Technological Adoption, places Guatemala in 43rd out of 140 countries. Given that it is an important pillar for business competitiveness, the following article focuses on the determination of technological trends and data analysis tools used by students and alumni of the Rafael Landívar University of Guatemala, because the current trends in Big Data and Small Data and its importance in business intelligence and business decision-making.

Keywords: trend, technology, big data, small data, data analysis, patterns, clues, technological tools.

Introducción

Este trabajo pretende profundizar en el conocimiento de las tendencias y tecnologías enfocadas en el análisis de datos en las carreras de la Facultad de Ciencias Económicas y Empresariales de la Universidad Rafael Landívar, y determinar la importancia de los mismos a través de un estudio de alcance descriptivo.

Se describen los temas: (a) tendencias tecnológicas, (b) tendencias en el análisis de datos y (c) simbiosis entre *Big Data* y *Small Data*. Además se detalla la importancia que tienen para cada una de las carreras de la Facultad de Ciencias Económicas y Empresariales: (a) las herramientas tecnológicas; (b) las técnicas de análisis de datos y (c) las tendencias tecnológicas relacionadas a datos. Los resultados reflejan la importancia del análisis de datos, como tendencia tecnológica y simbiosis o enlace entre el *Big Data* y *Small Data*.

1. Revisión literaria

¿Qué son las tendencias?

En forma general, según una de las definiciones que da el *Diccionario de la lengua española*, tendencia es una «idea religiosa, económica, política, artística, etc., que se orienta en determinada dirección»; lo que significa que las tendencias convergen en muchos ámbitos de la vida. En este siglo, el ámbito donde más uso presenta el término es en las redes sociales, bajo el concepto de *trending topics* o temas de tendencia; enfocados a los ámbitos de la vida.

¿Qué es la tecnología?

Según Coyle (2017), en su libro *El producto interno bruto*, una historia breve, pero entrañable, «La “tecnología” toma la forma de ideas en las mentes de las personas, o de educación y habilidades, o de ideas incorporadas en equipos y productos» (p. 107). Este concepto de tecnología lo enfoca como un tema de tendencia posicionado en la mente de las personas.

En su libro *Apuntes de crecimiento económico*, Sala i Martín (1999) explica que «por tecnología entendemos la “formula” o “conocimiento” que permite a las empresas “mezclar” capital y trabajo para producir un producto atractivo para los consumidores» (p. 167). Este concepto lo enfoca en el entorno empresarial como una tendencia a utilizar el conocimiento para atraer a los consumidores. Así mismo expresa que «una característica física importante que tiene la tecnología o la fórmula es que se trata de un bien “no rival” en el sentido que puede ser utilizada por mucha gente al mismo tiempo» (p. 168); describe una característica vital para que la tendencia se mantenga, ya que al ser un bien no rival, la misma se vuelve útil a muchas personas y la aceptación crece.

En relación con lo anterior, Sala i Martín (1999) advierte que no debe confundirse el concepto de rivalidad con el de exclusión:

El concepto de “rivalidad” no debe ser confundido con el concepto de capacidad de exclusión. Este segundo concepto se refiere a si podemos evitar que alguien utilice un determinado producto. Si se puede evitar se dice que el bien es excluible y si no, se dice que el bien es no excluible (p. 168).

Desde este punto de vista la exclusión en tecnología se enfoca desde el costo económico, las patentes o la creación del *know how* o conocimiento propio de la empresa.

De acuerdo con García, Alonso, y Fernández (2015), la tecnología es una gran palanca para lograr mejores resultados, pero por sí sola la tecnología no es la solución mágica para lograr las metas y objetivos, debe ser una herramienta al servicio del talento de las personas.

1.1. Tendencias tecnológicas

Según Ortiz (2014) la sociedad del siglo XX generó los avances necesarios para una sociedad que en el siglo XXI está en comunicación constante por medio de las redes sociales y con avances tecnológicos que mejoran las condiciones de vida de la sociedad.

La sociedad del siglo XX se caracterizó por el desarrollo de grandes inventos que le permitieron a la humanidad vivir con mayor comodidad, gozar de mejores condiciones de vida y comunicarse con mayor eficiencia, sin importar las distancias. El número de avances tecnológicos durante estos años, aumentan exponencialmente respecto al tiempo, hecho que ha generado, en este nuevo siglo, el nacimiento de otros rasgos culturales y artísticos debido a una rápida adaptación y aceptación del ser humano hacia los nuevos elementos tecnológicos y a su vez, una ruptura con el pasado (p. 51).

Aunque los avances tecnológicos aumentan exponencialmente respecto al tiempo, desde la percepción del ser humano estos avances tecnológicos se mantienen en tendencia por la aceptación que puedan presentar.

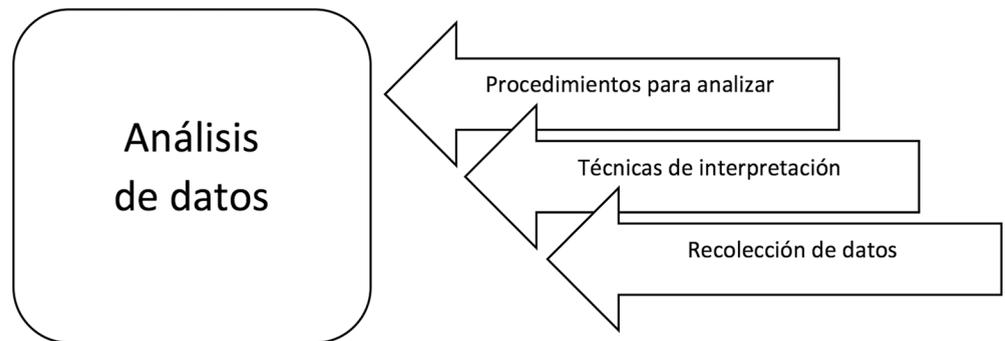
1.2. Tendencias en el análisis de datos

Análisis de datos, es un término que en los últimos años ha incrementado el interés de las personas, especialmente en el ámbito empresarial. Este incremento de interés surgió por las nuevas tendencias empresariales enfocadas en la inteligencia de negocios y en procesos de toma de decisiones basados en datos para lograr ventajas competitivas, pero en realidad el termino no es nuevo.

El estadista John Wild Tukey (1961) definió el análisis de datos en su artículo *The Future of Data Analysis*, de la siguiente forma: "Procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate" (p. 2) («Procedimientos para analizar datos, técnicas para interpretar los resultados de dichos procedimientos, formas de planificar la recopilación de datos para hacer su análisis más fácil, más preciso o más certero»).

El concepto que presenta Tukey es bastante completo, abarca todos los puntos clave del porqué el análisis de datos tiene un interés especial en la inteligencia de negocios y toma de decisiones basados en el objetivo de lograr precisión y asertividad, como se presenta en la figura 1.

Figura 1. Análisis de datos



Fuente: elaboración propia, 2018.

Según García, Alonso, y Fernández (2015), el análisis de datos es un proceso complejo de realizar y aconseja lo siguiente sobre el tema:

- Medir no es suficiente.
- No solo hay que medir y tener datos, hay que generar conocimiento sobre ellos e interiorizarlos. Medir no es igual a tener buenos datos. Es tan importante, si no más, saber qué tienes que medir, que el acto de medir.
- Desconocemos la mayoría de las métricas, medimos mucho, pero medimos mal, medimos, y no sabemos qué hacemos y con qué impacto.
- Grandes decisiones bien fundamentadas con datos no necesariamente tienen por qué ser las mejores o las que más éxito generen. La realidad es que no todo lo importante para decidir se puede medir. La experiencia y el alto entrenamiento generan capacidades muy desarrolladas que crean una especie de olfato, una intuición rica en detalles muy valiosa para decidir.
- Las diferencias duraderas se logran gracias a lo que no se puede copiar. Y eso suele estar vinculado con lo que hacen las personas, cómo aprenden y cómo aplican sus conocimientos.

La importancia del análisis de datos ha generado que las técnicas o procedimientos para analizar se desarrollen y las opciones de análisis aumenten. Varias de las técnicas nuevas se han vuelto tendencias de aprendizaje, por lo cual se describen a continuación las más relevantes y se recomienda un *software* de uso libre para su aprendizaje.

Técnica: Aprendizaje de máquina o *Machine Learning*

Descripción: Técnica que busca que los sistemas aprendan.

Software: Orange, <https://orange.biolab.si/>

Software: e1071 de R, <https://cran.r-project.org/web/packages/e1071/>

Técnica: Pronósticos o *Forecast*

Descripción: Técnica que determina comportamientos futuros para reducir la incertidumbre con base a datos históricos.

Software: Excel, <https://products.office.com/>

Software: forecast de R, <https://cran.r-project.org/web/packages/forecast/index.html>

Técnica: Minería de datos o *Data Mining*

Descripción: Técnica cuyo objetivo es descubrir patrones y tendencias.

Software: Tanagra, <http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html>

Software: Rattle de R, <https://rattle.togaware.com/>

Técnica: Modelos de series de tiempo o *Time Series Models*

Descripción: Técnica cuyo objetivo es determinar tendencias y comportamientos cíclicos y estacionales para la toma de decisiones.

Software: tseries de R, <https://cran.r-project.org/web/packages/tseries/index.html>

Técnica: Algoritmos o *Algorithms*

Descripción: Técnica para la creación de modelos.

Software: Python, <https://www.python.org/>

Software: R, <https://www.r-project.org/>

Técnica: Muestreos o *Sampling*

Descripción: Técnicas para la selección de muestras.

Software: R, hfgaleros@url.edu.gt, TMuestreo_1.0

1.3. Simbiosis entre Big Data y Small Data

Las tendencias de análisis de datos actuales se enfocan en los procesos, las formas, las interpretaciones y los resultados que los datos les puedan brindar para la toma de decisiones. Actualmente hay dos tendencias totalmente opuestas, que si partimos del concepto presentado por Sala i Martín (1999) sobre la tecnología, como la fórmula o conocimiento, se encuadran en tendencias tecnológicas para el análisis de datos; el *Big Data* y el *Small Data*. Las mismas se describirán a continuación y se enmarcarán en el concepto de doble diamante utilizado en el *Desing Thinking Process* para innovación.

¿Qué es el *Big Data*?

En su libro *Big Data para directivos*, Solana y Roca (2015) presentan la definición de *Big Data* como «la toma de decisiones o la prestación de servicios basada en el uso de flujo de datos digitales y la capacidad de procesarlos en tiempo real» (p. 11). Desde el punto de vista de Solana, el *Big Data* presenta una oportunidad para las empresas, para mejorar su gestión empresarial generando nuevas oportunidades de negocios o mejorar los servicios ya prestados para desarrollar la marca y recuperar el control del mercado.

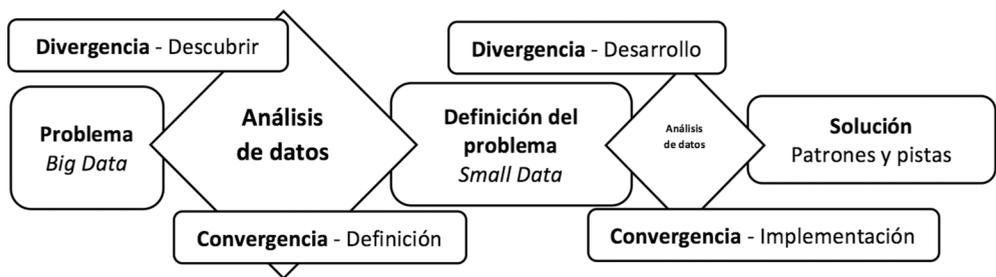
¿Qué es el *Small Data*?

En su libro *Small Data*, Lindstrom (2016) muestra pequeñas pistas que nos advierten de las grandes tendencias, y señala: «teniendo en cuenta que el equipo directivo no sabe qué hacer con el Big Data, todo el mundo está a la búsqueda del post-Big Data, y la respuesta son los pequeños datos» (p. 29). Desde su punto de vista, el *Small Data* puede brindar a la empresa las observaciones, las pistas, la conexión, la correlación, causalidad o concepto, que guíe a la organización a tomar la mejor decisión.

¿Qué es el *Desing Thinking Process*?

Es una forma de trabajo enfocada en la innovación en entornos inciertos de forma ágil y radical. Dentro de la forma de trabajo se encuentra el concepto de doble diamante utilizado para la resolución de problemas o toma de decisiones (Norman, 2013). La figura 2 muestra el concepto de doble diamante aplicado a *Big Data* y *Small Data* con base al análisis de datos.

Figura 2. Concepto de doble diamante aplicado a *Big Data* y *Small Data*



Fuente: elaboración propia, 2018.

2. Metodología de investigación

Se trata de un estudio de alcance descriptivo mediante encuesta, obtenido a partir de los datos procedentes de alumnos y exalumnos de la Universidad Rafael Landívar (26 al 28 de octubre de 2017). La encuesta fue enviada de forma masiva cuyo tamaño muestral resultante fue de 126 individuos.

Con relación al tamaño de muestra mínimo para realizar un estudio de alcance descriptivo, Sampieri (2014), recomienda 30 casos por grupo o segmento del universo.

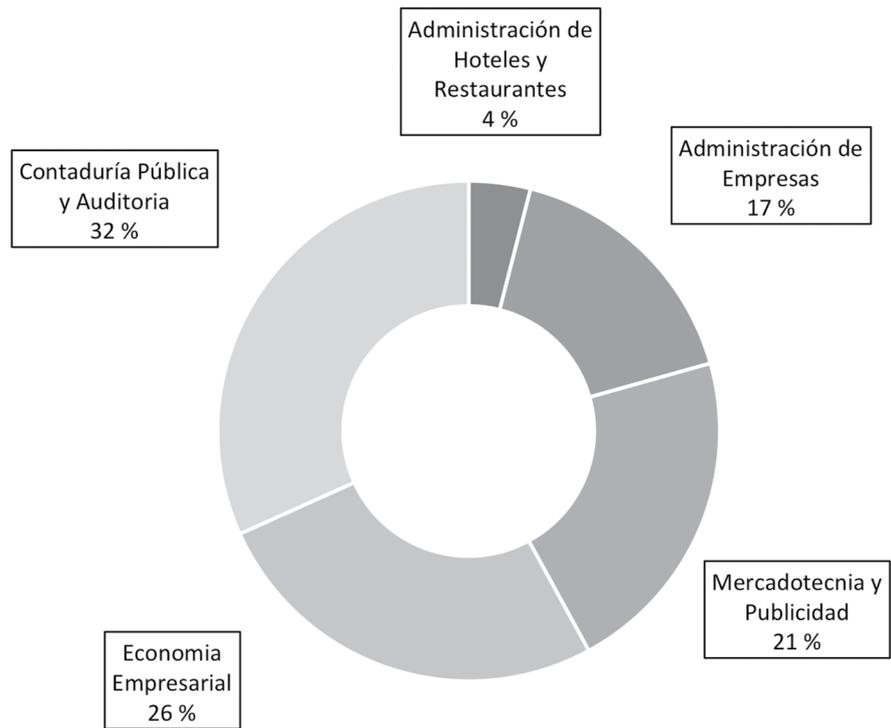
3. Resultados

A continuación se presentan los resultados obtenidos de la muestra evaluada.

Del total de encuestados (126 alumnos y exalumnos), el 44 % eran graduados de la carrera de licenciatura y el 56 % no habían completado aún la carrera.

La distribución de los porcentajes de las carreras de licenciatura se presenta en el siguiente gráfico.

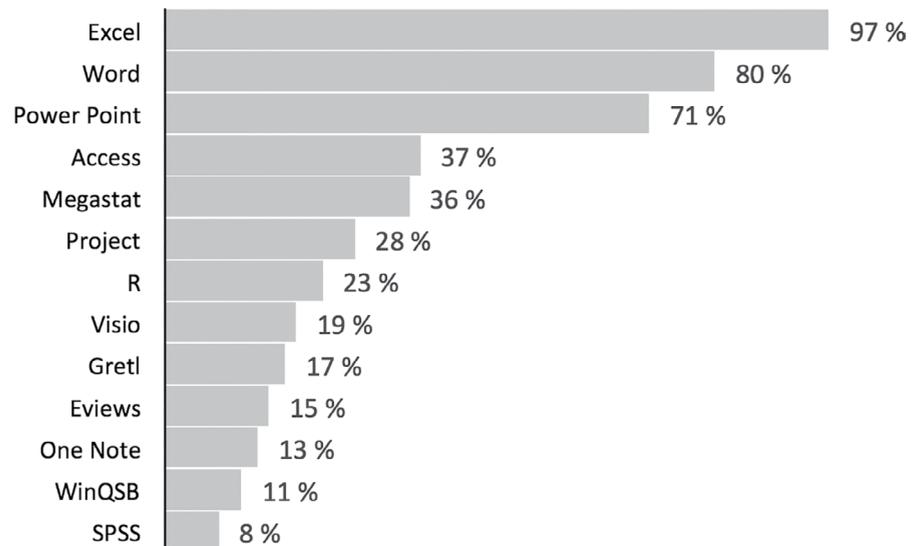
Gráfico 1. Carrera de licenciatura



Fuente: elaboración propia, 2018.

Con relación a la importancia de las herramientas tecnológicas, el 97 % indicó que Excel era una herramienta importante, pero además incluyó a Word y Power Point como herramientas predominantes, con un 80 % y un 71 % respectivamente.

Gráfico 2. Importancia de las herramientas tecnológicas



Fuente: elaboración propia, 2018.

Al partir de la importancia que representa la utilización de los paquetes de Office para los encuestados en la tabla 1, se presentan los porcentajes de importancia detallados por herramienta y carrera, con los siguientes resultados: en Administración de Empresas, Project es la herramienta más importante con un 8.5 %; en Administración de Hoteles y Restaurantes, Access, Visio y R presentan un 9.1 %; para Contaduría Pública y Auditoría la herramienta que presenta más relevancia después de Excel, Word y Power Point es Access, con un 13.2 %; para Economía Empresarial la herramienta Eviews presenta un 14.1 %; y en el caso de Mercadotecnia y Publicidad la herramienta educativa Megastat presenta un 9.3 %. El desglose se presenta en la tabla 1.

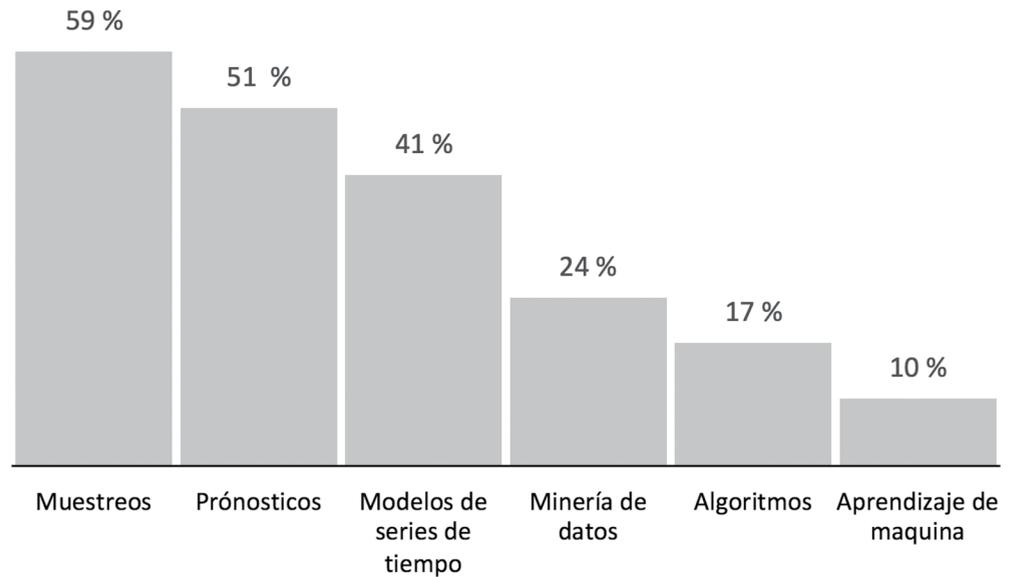
Tabla 1. Importancia de las herramientas tecnológicas por carrera

Herramienta	Administración de Empresas	Administración de Hoteles y Restaurantes	Contaduría Pública y Auditoría	Economía Empresarial	Mercadotecnia y Publicidad
Word	18.9 %	27.3 %	20.1 %	14.0 %	16.8 %
Excel	19.8 %	36.3 %	23.0 %	19.5 %	21.0 %
Power Point	17.9 %	9.1 %	15.5 %	12.8 %	17.6 %
Access	6.6 %	9.1 %	13.2 %	4.3 %	7.6 %
One Note	5.7 %	0.0 %	3.4 %	1.2 %	2.5 %
Project	8.5 %	0.0 %	5.2 %	4.3 %	8.4 %
Visio	4.7 %	9.1 %	4.0 %	1.2 %	7.6 %
Megastat	6.6 %	0.0 %	8.0 %	7.9 %	9.3 %
WinQSB	5.7 %	0.0 %	2.3 %	0.0 %	3.4 %
SPSS	0.0 %	0.0 %	1.2 %	8.5 %	2.5 %
R	1.9 %	9.1 %	1.7 %	8.5 %	1.7 %
Eviews	2.8 %	0.0 %	1.2 %	14.1 %	0.8 %
Gretl	0.9 %	0.0 %	1.2 %	3.7 %	0.8 %
TOTAL	100 %	100 %	100 %	100 %	100 %

Fuente: elaboración propia, 2018.

Desde el punto de vista de las técnicas de análisis de datos, la importancia se enfoca en los muestreos con un 59 %, los resultados completos se presentan en el gráfico 3.

Gráfico 3. Importancia de las técnicas de análisis de datos



Fuente: elaboración propia, 2018.

En importancia general, los muestreos presentaron el porcentaje más alto desde el punto de vista de cada carrera. Administración de Empresas presentó a los Modelos de series de tiempo con un 29 %; en el caso de la Administración de Hoteles y Restaurantes la importancia está por encima de los Muestreos y se enfoca en los Pronósticos con un 50 %; en el caso de Contaduría Pública y Auditoría, por debajo de los Muestreos la técnica más importante es la de los Pronósticos, igual que en el caso anterior. Por otra parte; en el caso de la carrera de Economía Empresarial, la importancia radica en las técnicas de Modelos de series de tiempo con un 28 %; y en el caso de Mercadotecnia y Publicidad en las técnicas de Pronósticos con un 32 % por encima de las técnicas de Muestreo. En la tabla 2 se presentan los resultados completos.

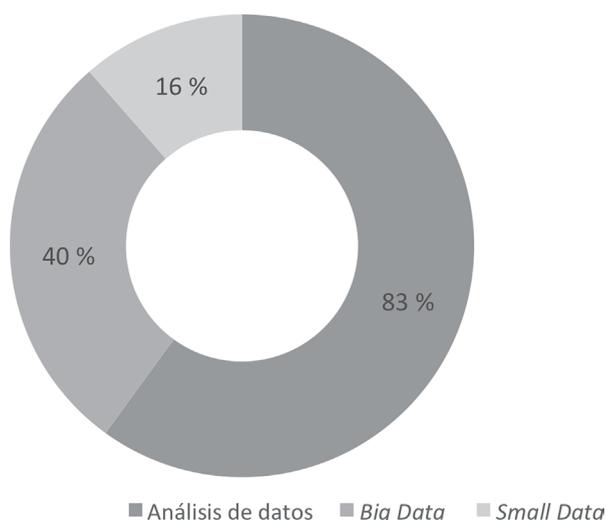
Tabla 2. Importancia de las técnicas de análisis de datos por carrera

Técnica	Administración de Empresas	Administración de Hoteles y Restaurantes	Contaduría Pública y Auditoría	Economía Empresarial	Mercadotecnia y Publicidad
Aprendizaje de máquina	11 %	0 %	2 %	6 %	3 %
Pronósticos	14 %	50 %	21 %	26 %	32 %
Minería de datos	7 %	10 %	16 %	13 %	7 %
Modelos de series de tiempo	29 %	0 %	14 %	28 %	19 %
Algoritmos	3 %	10 %	9 %	9 %	11 %
Muestreos	36 %	30 %	38 %	18 %	28 %
TOTAL	100 %	100 %	100 %	100 %	100 %

Fuente: elaboración propia, 2018.

Desde el punto de vista de las tendencias tecnológicas que los encuestados han manejado o desarrollado, se encuentran en orden de importancia: el Análisis de datos con un 83 %, el *Big Data* con un 40 % y el *Small Data* con un 16 %. En el gráfico 4 se presentan los resultados completos.

Gráfico 4. Importancia de las tendencias tecnológicas relacionadas a datos



Fuente: elaboración propia, 2018.

Desde el enfoque de las carreras, el Análisis de datos es la que con mayor porcentaje refleja en cada una de las tendencias. En segundo plano el *Big Data* se presenta con mayor porcentaje que el *Small Data*. En la tabla 3 se presentan los resultados completos.

Tabla 3. Importancia de las tendencias tecnológicas relacionadas a datos por carrera.

Tendencia	Administración de Empresas	Administración de Hoteles y Restaurantes	Contaduría Pública y Auditoría	Economía Empresarial	Mercadotecnia y Publicidad
Big Data	26 %	17 %	24 %	38 %	26 %
Small Data	4 %	0 %	11 %	17 %	11 %
Análisis de datos	70 %	83 %	65 %	45 %	63 %
TOTAL	100 %	100 %	100 %	100 %	100 %

Fuente: elaboración propia, 2018.

Conclusiones

Al trasladar las tendencias tecnológicas aplicadas a datos al proceso de doble diamante utilizado en *Design Thinking* para innovación, es evidente que el Análisis de datos es la parte medular y simbiótica del proceso donde está involucrado el *Big Data* y el *Small Data*.

Al considerar la importancia de las herramientas tecnológicas, los resultados muestran que los paquetes como Word, Excel, Power Point y Access de Microsoft Office juegan un papel vital y de alta importancia en cualquiera de las carreras al momento de su desarrollo y de la aplicación de los conocimientos en el ámbito laboral.

En todas las carreras, sobre todo para Administración y Contaduría Pública y Auditoría el aspecto más importante en las técnicas que apoyan al Análisis de datos, es el muestreo.

Por su parte, en las tendencias tecnológicas enfocadas a datos, el Análisis de datos y el *Big Data* son las que cuentan con mayor importancia dentro de las carreras de la Facultad de Ciencias Económicas. El tema de *Small Data* inicia a colocarse como un tema de tendencia, pero no se ha desarrollado por completo.

Referencias

- Coyle, D. (2017). *El producto interno bruto*. México D. F., México: Fondo de cultura económica.
- García, J.; Alonso, A. y Fernández, A. (2015). *Nunca te fies de un economista que no dude*. Barcelona, España: Deusto.
- Lindstrom, M. (2016). *Small Data. Las pequeñas pistas que nos advierten de las grandes tendencias*. Barcelona, España: Deusto.
- Norman, D. (2013). *The design of everyday things*. New York, United States: Basic Books.
- Ortiz, V. A. (2014). *Ideas metodologicas para enseñar y aprender matematica*. Cara Parens.
- Real Academia Española. (2018). *Diccionario de la lengua española*. Disponible en: <https://dle.rae.es/?id=ZSZtERx>
- Sala i Martin, X. (1999). *Apuntes de crecimiento económico*. Barcelona, España: Antoni Bosch.
- Sampieri, R. H. (2014). *Metodologias de la investigacion*. México D. F., México: Mc Graw Hill Education.
- Solana, A. y Roca, G. (2015). *Big Data para directivos*. Barcelona, España: Ediciones Urano.
- Tukey, J. W. (1961). The future of data analysis. *The Annals of Mathematical Statistics*, 33(1), 1-67.