



Organización
Internacional
del Trabajo

CINTERFOR



Lineamientos prácticos para el uso del *Big Data* e identificación de brechas laborales

Yesica Fernández Malo
Fernando Vargas Zúñiga

Copyright © Organización Internacional del Trabajo (OIT/Cinterfor) 2022

Las publicaciones de la Oficina Internacional del Trabajo gozan de la protección de los derechos de propiedad intelectual en virtud del protocolo 2 anexo a la Convención Universal sobre Derecho de Autor. No obstante, ciertos extractos breves de estas publicaciones pueden reproducirse sin autorización, con la condición de que se mencione la fuente. Para obtener los derechos de reproducción o de traducción deben formularse las correspondientes solicitudes a Publicaciones de la OIT (Derechos de autor y licencias), Oficina Internacional del Trabajo, CH - 1211 Ginebra 22, Suiza, o por correo electrónico a rights@ilo.org, solicitudes que serán bien acogidas.

Las bibliotecas, instituciones y otros usuarios registrados ante una organización de derechos de reproducción pueden hacer copias de acuerdo con las licencias que se les hayan expedido con ese fin. En www.ifrro.org puede encontrar la organización de derechos de reproducción de su país.

Fernández Malo, Yesica; Vargas Zúñiga, Fernando

Lineamientos prácticos para el uso del Big data e identificación de brechas salariales.
Montevideo: OIT/Cinterfor, 2022, 56 páginas.

ISBN: 978-92-9088-306-7

Las denominaciones empleadas, en concordancia con la práctica seguida en las Naciones Unidas, y la forma en que aparecen presentados los datos en las publicaciones de la OIT no implican juicio alguno por parte de la Oficina Internacional del Trabajo sobre la condición jurídica de ninguno de los países, zonas o territorios citados o de sus autoridades, ni respecto de la delimitación de sus fronteras.

La responsabilidad de las opiniones expresadas en los artículos, estudios y otras colaboraciones firmadas incumbe exclusivamente a sus autores, y su publicación no significa que la OIT las avale.

Las referencias a firmas o a procesos o productos comerciales no implican aprobación alguna por la Oficina Internacional del Trabajo, y el hecho de que no se mencionen firmas o procesos o productos comerciales no implica desaprobación alguna.

Para más información sobre las publicaciones y los productos digitales de la OIT, visite nuestro sitio web: ilo.org/publns.

Para más información sobre esta publicación, contáctese con el Centro Interamericano para el Desarrollo del Conocimiento en la Formación Profesional (OIT/Cinterfor), Av. Uruguay 1238, Montevideo, Uruguay. Visite nuestro sitio <http://www.oitcinterfor.org> o escríbanos a oitcinterfor@ilo.org.

OIT/Cinterfor es un servicio técnico de la OIT, establecido en 1963 con el fin de impulsar y coordinar los esfuerzos de las instituciones y organismos dedicados a la formación profesional en la región.

Esta publicación se terminó de imprimir en el Departamento de Publicaciones de OIT/Cinterfor, en la ciudad de Montevideo, en el mes de setiembre 2022

Impreso en Uruguay

▶ Tabla de contenido

Presentación	6
Introducción	8
1. Desajustes y brechas de habilidades	10
1.1 Medición tradicional de las brechas de habilidades	12
2. El uso del <i>Big Data</i> en la medición de brechas de habilidades.....	16
2.1. ¿Qué es el Big Data?.....	16
2.2. Conceptos claves: Disciplinas y el Big Data	17
2.2.1. Minería de datos <i>Data Mining</i>	18
2.2.2. Aprendizaje automático <i>Machine Learning</i>	20
2.2.3. Minería web <i>Web Mining</i>	22
3. Usando el Big Data para procesar las Ofertas de Trabajo en Línea (OTL)	24
3.2. Pasos claves para el procesamiento de Bases de Datos	30
3.3. Pasos claves para el procesamiento de Bases de Datos	32
3.3.1. Uso de algoritmos	33
3.3.2. Uso de datos	34
3.3.3. Reutilización de datos y nuevos tipos de datos.....	34
3.3.4. Taxonomía de los datos	35
3.4. Experiencias internacionales del <i>Big Data</i> en la identificación de brechas de habilidades y de talento humano	36
3.4.1. Sector público.....	36
3.4.2. Sector privado	40
3.4.3. Casos internacionales	44
Conclusiones	48
Anexos.....	50
Referencias	52

▶ Lista de figuras

Figura 1. Consecuencias de las brechas de Capital Humano	12
Figura 2. Evolución de servicios de intermediación laboral.....	14
Figura 3. Disciplinas y el <i>Big Data</i>	17
Figura 4. Proceso general de la minería de datos.....	20
Figura 5. Algunos usos actuales del <i>Machine Learning</i>	21
Figura 6. Principales características de una red de expertos para desarrollar <i>Big Data</i> para IML (Información del Mercado Laboral)	26
Figura 7. El proceso DCBD y las grandes “V” de <i>Big Data</i> involucradas en cada paso	31
Figura 8. Parámetros estadísticos de <i>Big Data</i>	33
Figura 9. Sitio web Skills-OVATE	37
Figura 10. Sitio web Internet Vacancy Index	38
Figura 11. Sitio web Current Board Vacancies	39
Figura 12. Sitio web Cyber Seek	40
Figura 13. Sitio web The iLabourProject	41
Figura 14. Sitio web Burning Glass Technologies.....	42
Figura 15. Sitio web Bizkaia Talent	43
Figura 16. Sitio web Monitor Laboral	45
Figura 17. Estudio de análisis de oferta y demanda de empleo, República Dominicana.....	46
Figura 18. Estudio de análisis de oferta y demanda de empleo, Uruguay.....	

▶ Lista de tablas

Tabla 1. Tipo de desajuste de habilidades.....	12
Tabla 2. Indicadores y fuentes para la medición de brechas.....	13
Tabla 3. Ventajas y desventajas de OTL.....	25
Tabla 4. Habilidades más demandadas resultado de <i>webscrapig</i> para Paraguay 2021.....	28

▶ Presentación

Uno de los principales motores para el desarrollo económico y social de un territorio es su capital humano, la importancia de realizar un seguimiento a los datos y registros administrativos que permitan diagnosticar las problemáticas que afectan las dinámicas laborales, como las barreras que impiden tener altos factores de competitividad, productividad y empleabilidad; lleva a la necesidad de contar con metodologías y lineamientos prácticos, que permitan medir de manera ágil las brechas de habilidades, como poder identificar las ocupaciones críticas para crear acciones y políticas que permitan su cierre.

Este documento busca brindar lineamientos prácticos para facilitar la comprensión, acompañamiento y aplicación de técnicas de análisis masivo de datos (*Big Data*) que permita la toma de decisiones fundamentada para implementar políticas de análisis, medición y cierre de brechas de talento humano.

Los datos en línea sobre los mercados laborales se han convertido en los últimos años en una importante fuente de información para comprender mejor cómo funcionan los mercados laborales. Este proceso se ha visto afectado por la expansión de Internet y la aparición de plataformas intermediarias del mercado laboral en línea. Aunque la gran totalidad de países disponen de encuestas de hogares y empleo, que permiten un seguimiento constante a la fuerza laboral, tienen problemáticas recurrentes de falta de representatividad estadística a nivel de ocupación y cargo, las informaciones pueden no estar disponibles en tiempo cortos, y en ocasiones no cubren aspectos tan específicos de los mercados laborales con suficiente detalle y frecuencia, como por ejemplo una visión de las habilidades o competencias con las que cuenta la oferta laboral. La ausencia de datos de encuestas detallados y de alta frecuencia y la creciente disponibilidad de datos

en línea, ha impulsado el crecimiento de investigaciones basadas en datos en línea para estudiar diversas preguntas.

Este documento explicará de manera práctica cómo, a partir de procesos de *Big Data* o uso de datos en línea es posible avanzar en el análisis de la fuerza laboral, demanda laboral y oferta educativa. El objetivo es mostrar por medio de ejemplos prácticos y experiencias internacionales el proceso general y los pasos claves para incentivar el uso del *Big data* para el análisis del Mercado laboral.

No se soslayan algunos factores críticos marcados en la literatura internacional; los datos suelen carecer de estructura, pueden presentar duplicaciones y se argumenta en muchos casos su desafío de representatividad. Normalmente necesitan “limpieza” y verificación de calidad, y están sujetos a muchos otros problemas potenciales, incluidos los relativos a la privacidad que tienden a delimitar su uso efectivo.

En varios países de la región, una limitación adicional es el escaso uso de las vacantes en línea debido a la mala conectividad en regiones periféricas y una gran proporción de trabajos informales. Sin embargo, las ofertas de empleo en línea y otros tipos de análisis de *Big data* tienen un gran potencial para contribuir a una mejor comprensión de los mercados laborales, especialmente si se complementan con fuentes de información más tradicionales, como encuestas de hogares o empleo, y registros administrativos de los observatorios de trabajo.

Este informe ha sido preparado por Yesica Fernández, Consultora de OIT/Cinterfor con el apoyo y orientación de Fernando Vargas, Especialista Senior de OIT/Cinterfor.

► Introducción

Los mercados laborales son dinámicos, cambian constantemente y por tanto, se enfrentan a cambios inesperados que afectan su comportamiento. Es por esto que, es un desafío importante para los hacedores de política pública, lograr identificar las necesidades del mercado para traducirlas en requerimientos de habilidades y competencias, de tal manera que se detecten los desbalances en la dinámica laboral producidos en diversas vías, bien sea desde las empresas que no logran encontrar las habilidades que necesitan o, desde las personas porque no poseen las habilidades requeridas y no tienen suficiente información sobre qué habilidades desarrollar para conseguir un empleo.

Bajo este contexto, los formuladores de políticas laborales cuentan con informaciones como los datos de encuestas nacionales de hogares y empleo, las cuales a pesar de ser fuentes de información válida y útil, pueden ocasionalmente presentar limitaciones como la falta de regularidad en la aplicación de la encuesta o un déficit en la captura de información específica que facilite acciones de manera completa y confiable¹. Además, no permiten evidenciar claramente las necesidades de habilidades asociadas con ocupaciones particulares que varían con el contexto y cambian con el tiempo.

Según la Organización Internacional del Trabajo “las fuentes emergentes de datos sobre habilidades tienen el potencial de proporcionar información detallada y en tiempo real sobre las necesidades de una manera rentable”. Gracias a los avances tecnológicos, la digitalización y las plataformas de Internet se han dado avances para recopilar conjuntos de datos muy grandes y ricos reconocidos como *Big Data*, los cuales sirven para muchos propósitos.²

Los datos en línea sobre los mercados laborales se han convertido en los últimos años en una importante fuente de información para comprender

1 ILO, 2020. *The feasibility of using big data in anticipating and matching skills needs*. Geneva, 2020 https://www.ilo.org/skills/areas/skills-training-for-poverty-reduction/WCMS_759330/lang--en/index.htm

2 ILO, 2020.

mejor cómo funcionan los mercados laborales. Este proceso se ha visto afectado por la expansión de Internet y la aparición de plataformas intermediarias del mercado laboral en línea (por ejemplo, Babajobs en India, Glassdoor en los Estados Unidos, Profesia en Eslovaquia, Computrabajo en América Latina y sitios web profesionales y redes sociales como LinkedIn, Twitter o Facebook). Estos tipos de datos del mercado laboral actualmente brindan una fuente de datos oportuna, granulares y, a menudo, tan completos que los encargados de formular políticas utilizan cada vez más para analizar los mercados laborales en todo el mundo.

Por su parte, las Ofertas de Trabajo en Línea (OTL) contribuyen con la generación de enormes conjuntos de datos que contienen información detallada sobre los requisitos anunciados. Esta información generalmente registra características más amplias sobre la vacante, tales como: habilidades requeridas, títulos de los puestos, requisitos de calificaciones, certificaciones y experiencia, datos del empleador, sector económico, categoría ocupacional y la ubicación geográfica del puesto anunciado. Estos datos de ofertas de empleo en línea están disponibles en tiempo real y, a diferencia de los datos de las encuestas que requieren tiempo de procesamiento antes de la publicación, se pueden utilizar casi de inmediato, ofreciendo una ventaja sobre las taxonomías de requisitos de habilidades, que generalmente requieren una cantidad considerable de investigación y análisis, junto con tiempo para ser producidas y actualizadas regularmente.

Así, este documento se divide en tres capítulos que permitirán conocer las principales técnicas de *Big Data* para la lectura y seguimiento del mercado laboral. A partir de la contextualización del capital humano y la identificación de ocupaciones críticas, el **primer capítulo** ofrece herramientas para entender y dimensionar las principales brechas de talento humano expresadas por ocupaciones. El **segundo capítulo** presenta los aspectos conceptuales del *Big Data* y el **tercer capítulo** centra su análisis en el uso del *Big Data* para el análisis de mercado laboral evidenciando las diferencias prácticas entre su uso y la estadística tradicional, en el uso de la herramienta para identificar las brechas de habilidades. Por último, se presentan algunas experiencias del manejo del *Big Data* en la identificación de brechas del capital humano.

1 ▶ Desajustes y brechas de habilidades

La brecha de habilidades mide el grado en que los trabajadores carecen de las habilidades necesarias para desempeñar alguna actividad, ello está usualmente determinado por la percepción de los empleadores sobre las deficiencias de habilidades del trabajador (ILO, 2019). En este orden, la brecha de habilidades laborales puede definirse como el desajuste entre las habilidades demandadas por los sectores productivos y las habilidades con las que cuenta la fuerza de trabajo.

En la Tabla 1. Tipo de desajuste de habilidades, se explican los diferentes tipos de desajustes y brechas de habilidades:

Tabla 1. Tipo de desajuste de habilidades

Brecha de habilidades (<i>Skills gap</i>)	El trabajador no dispone de las habilidades necesarias para realizar su trabajo actual. Se estima a partir de la opinión de las empresas sobre las deficiencias de habilidades que perciben entre sus trabajadores
Escases de habilidades	La demanda de trabajadores con una habilidad particular excede la oferta de trabajadores disponibles con esa habilidad. Se suele medir a través de las vacantes no cubiertas o de difícil ocupación, debido a la falta de candidatos para el puesto.
Sobreoferta de habilidades	La oferta de trabajadores con una habilidad particular excede la demanda de trabajadores con esa habilidad.
Desajuste o brecha vertical de habilidades	
(Sub/sobre) cualificación (<i>Undereducation o overeducation</i>)	Los trabajadores poseen un nivel educativo menor/superior del requerido para realizar el trabajo actual
(sub/sobre) capacitación (<i>Underskilling o overskilling</i>)	Los trabajadores carecen/exceden del tipo de y nivel habilidades necesarias para realizar su trabajo actual
Desajuste o brecha horizontal de habilidades	
Desajuste o brecha por campo de estudio (<i>Field of study mismatch</i>)	El trabajador (por lo general, graduado universitario) tiene una ocupación que no está relacionada con su campo de estudio
Obsolescencia de habilidades	Las habilidades que se necesitan en un trabajo ya no se requieren o no son tan importantes, o el trabajador las ha perdido

Tomado de: ILO, 2019. Skills and Jobs Mismatches in Low and Middle-income Countries.

La demanda por habilidades está cambiando debido al rápido avance tecnológico, la automatización y la digitalización, así como por otros factores como la globalización, la demografía y la transición ambiental. El desajuste o brecha de habilidades de la fuerza laboral, representa un obstáculo para el crecimiento económico. Estas brechas generan consecuencias como la baja productividad, el desempleo, la escasa participación laboral y la informalidad. La capacidad de responder a la demanda cambiante de habilidades se considera clave para las transiciones económicas exitosas que inevitablemente buscan las economías y las personas en todo el mundo, y para la recuperación del desarrollo entre los países de ingresos altos, medios y bajos³ (OIT, 2016).

Los niveles de productividad y competitividad de un país dependen, en gran parte, del capital humano con el que cuenta y, por tanto, disponer de trabajadores altamente calificados, suficientes y con competencias relevantes a la vocación productiva local es posiblemente uno de los factores que en mayor medida diferencia a los países más desarrollados del resto. De la misma manera, su ausencia representa uno de los principales obstáculos para la sofisticación y diversificación de una economía.

Por consiguiente, entender las causas por las cuáles se genera esta brecha y conocer posibles soluciones resulta relevante dada la existencia de una gran brecha y su persistencia . Los desajustes de habilidades tienen implicaciones negativas porque no sólo afecta la posibilidad de que las personas encuentren un trabajo decente, sino también, porque es un freno para el crecimiento del empleo, el aumento en la productividad y el mejoramiento en términos de desarrollo (OIT, Vol. 4, 2017).

Desde la perspectiva de demanda laboral, pueden darse situaciones como la falta de gestión del talento humano por competencias y la baja apertura de vacantes porque no se consigue el personal con las competencias adecuadas; desde la oferta laboral, los trabajadores evidencian que no cuentan con el perfil que requieren las empresas para desempeñar un empleo y; desde la oferta educativa, se nota que no se está formando a las personas en lo que requiere el sector productivo, según la -OIT- “si estas brechas no se identifican ni se miden previamente, no es posible formular políticas para su cierre” (OIT - Vol.2, 2017).

3 Compendio de guías metodológicas sobre anticipación y ajuste entre oferta y demanda de habilidades, publicado en: <https://www.oitcinterfor.org/node/7070>

Figura 1. Consecuencias de las brechas de capital humano



Fuente: Elaboración propia a partir de (OIT, 2017).

Adecuar esta oferta y demanda de competencias y habilidades, coadyuva a la disminución del desempleo (sobre todo el juvenil) y permite mejorar la calidad de vida de las personas a través de la promoción de la empleabilidad, la movilidad social y la inclusión (OIT - Vol.3, 2017).

1.1. Medición tradicional de las brechas de habilidades

Para establecer estrategias de cierre de brechas de capital humano, se debe contar previamente con metodologías que faciliten la identificación y medición. Esto requiere de ejercicios que permitan comprender el tipo de brecha de capital humano y por qué es importante avanzar en diseñar orientaciones precisas para su adecuada medición.

Una brecha de capital humano puede entenderse como la “grieta” o “fisura” que existe cuando la oferta y la demanda de trabajo no pueden encontrar un punto de encuentro o de equilibrio en el mercado de trabajo. Por consiguiente, estas brechas pueden ser consideradas como fallas de coordinación en la dinámica laboral cuando la oferta y la demanda no logran encontrarse entre sí.

Teniendo en cuenta los tipos de desajustes de habilidades y sus diferencias (Gontero & Novella, 2021) es posible plantear una serie de indicadores y métodos para realizar un ejercicio de medición de brechas como se muestra en la tabla a continuación.

Tabla 2. Indicadores y fuentes para la medición de brechas de habilidades

Tipo de brecha	Concepto	Tipo de análisis	Fuentes de información	Limitaciones de los indicadores
Macro-económicas	Miden la dispersión relativa de la tasa de ocupación entre los niveles educativos de la población	Cuantitativo	Encuestas de Hogares.	<ul style="list-style-type: none"> ▶ Pueden existir restricciones de las fuentes de información. ▶ No permite identificar si los trabajadores están trabajando en puestos de trabajo acorde a sus cualificaciones.
Perfilamiento	Ocurre cuando los perfiles de los oferentes no se ajustan a los requisitos de las vacantes que publican los empleadores	Cualitativo	Estudios sectoriales.	▶ La información solo se encuentra disponible si se ha realizado algún estudio, si no, se requiere el levantamiento de la información.
		Cuantitativo	Base de datos de vacantes y oferentes u Ofertas de empleo en línea.	▶ Puede que la información de las bases no esté estandarizada.
Competencias	Cuando las competencias que tienen los oferentes de empleo son diferentes a las que requiere el sector productivo	Cualitativo	Encuestas de escasez de talento humano.	▶ Falta de estrategias para combinar resultado de entrevistas con el sector productivo.
		Cuantitativo	Bases de datos de vacantes y oferentes. Entrevistas a empresas y gremios y a instituciones educativas.	<ul style="list-style-type: none"> ▶ Las encuestas de hogares no ofrecen información robusta de oferta laboral. ▶ La información solo se encuentra disponible solo si se ha realizado algún estudio, si no, se requiere el levantamiento de la información.
Ajuste Temporal	Hacen referencia al tiempo en que se demora un buscador de empleo en conseguir trabajo o una vacante	Cuantitativo	Bases de datos de vacantes y oferentes. Encuestas de Hogares.	▶ No se logra tener una medición específica por cargo y ocupación debido a que la información puede no ser representativa o actualizada.
Pertinencia	Se refieren a aquellas asociadas a la falta de participación del sector productivo en las etapas de diseño, planeación y ejecución de la educación terciaria	Cualitativo	Entrevistas a empresas y gremios y a instituciones educativas.	<ul style="list-style-type: none"> ▶ Objetividad y veracidad con que sea respondida la pregunta. ▶ Falta de documentación de programas por competencias.
Calidad	Se refieren a la al nivel de logro de competencias genéricas y específicas por parte de la oferta laboral	Cualitativo	Para demanda laboral: Estudios y entrevistas a empresas, centros de empleo. Para oferta educativa: Bases de datos de Instituciones de Educación Superior (IES), o Ministerios de Educación, entrevistas a IES, pensum y mallas curriculares.	<ul style="list-style-type: none"> ▶ La información solo se encuentra disponible solo si se ha realizado algún estudio, si no, se requiere el levantamiento de la información. ▶ Deficiencias en la información de oferta educativa.

Tipo de brecha	Concepto	Tipo de análisis	Fuentes de información	Limitaciones de los indicadores
Cantidad	Asociadas con el déficit (y/o ausencia) de oferta de programas de formación, con el déficit de demanda por programas de formación y con la baja capacidad de atracción (o retención) de capital humano relevante por parte de un sector específico	Cuantitativo	Encuestas de Hogares. Bases de datos de vacantes y oferentes. Sistemas de información de graduados por nivel de formación.	<ul style="list-style-type: none"> ▶ Las encuestas de hogares no ofrecen información robusta de oferta laboral. ▶ Es posible que no se cuente con información actualizada de los graduados.

Fuente: Elaboración propia a partir de OIT, 2017. Metodología de identificación y medición de brechas de capital humano -IMBCH.

Los requerimientos de las empresas en la búsqueda de personal con las habilidades y competencias necesarias para cubrir sus vacantes sumado a la evolución constante de herramientas tecnológicas para su registro, han construido una nube extensa de información expresada en variables y datos de la oferta disponible en el mercado laboral. Los servicios de empleo, o portales de empleo desempeñan un papel fundamental en la organización de la demanda y la oferta laboral, ya que son un punto de encuentro a través de la difusión de información sobre el mercado y la asistencia en la búsqueda de empleo y servicios de colocación.

De acuerdo con OIT/Cinterfor (Vargas, 2022) en los primeros años de este tipo de servicios, la información era registrada a mano en formularios extensos sobre los datos de las vacantes que las empresas demandaban y a su vez, sobre las necesidades de las personas disponibles para trabajar. Por lo que estos métodos eran insuficientes para procesar altas cantidades de información y de forma periódica que fuera capaz de retroalimentar procesos más allá del registro de oferentes y la tasa de colocación.

Los servicios de empleo en los años 90 automatizaron el *match* o el encuentro entre la oferta y la demanda, a partir de la comparación de registros de oferentes contra vacantes según el criterio del intermediario laboral, de acuerdo con uso de las clasificaciones de ocupaciones, que figuraba como único recurso para encontrar la pareja ideal (el *match*) para cada vacante. Por el ritmo de avance de la tecnología, los servicios de empleo tuvieron que adaptarse a procesos de transformación, que pasan desde la eliminación de procesos manuales hasta la incorporación de *Big Data*, Inteligencia Artificial,

robotización, ciencia de datos, el aprendizaje automático y el internet de las cosas (Urquidi, Ortega, Arza, & Ortega, 2021).

Figura 2. Evolución de servicios de intermediación laboral



Fuente: Elaboración propia a partir de Vargas, 2022.

La tendencia actual tiene que ver con la utilización de herramientas de análisis masivo de datos (*Big Data*) que incluyen elementos de aprendizaje de máquinas e inteligencia artificial; para el análisis del mercado laboral. Dentro de los temas que se analizan se encuentran:

- ▶ Ocupaciones y habilidades más requeridas demandas
- ▶ Ocupaciones disponibles en la oferta laboral con más frecuencia.
- ▶ Habilidades futuras y ocupaciones emergentes.
- ▶ Nuevas competencias digitales.
- ▶ Habilidades transversales o blandas.
- ▶ Ocupaciones críticas por cambios en el empleo.
- ▶ Desbalances entre oferta y demanda

2 ▶ El uso del *Big Data* en la medición de brechas de habilidades

2.1. ¿Qué es el *Big Data*?

Si bien no existe una definición única, puede ser entendido como un fenómeno de datos que surgen de la interacción espontánea con dispositivos interconectados (ICO, 2018). Estos a su vez, son estudiados y procesados por métodos computacionales y numéricos, debido a las características comunes que estos presentan (Sosa Escudero, 2021).

El uso de *Big Data* se ha convertido en capital para las empresas que utilizan información para mejorar la toma de decisiones, introducir nuevos mercados y ofrecer mejores experiencias de cliente. Entre otras definiciones se resaltan las siguientes:

- ▶ “Conjuntos de datos cuyo tamaño va más allá de la capacidad de captura, almacenado, gestión y análisis de las herramientas de base de datos tradicionales” (McKinsey & Company, 2011).
- ▶ “Volumen masivo de datos, tanto estructurados como no-estructurados, los cuales son demasiado grandes y difíciles de procesar con las bases de datos y el software tradicionales” (DNP, 2016).
- ▶ “*Big Data* son activos de información caracterizados por su alto volumen, velocidad y variedad, que demandan soluciones innovadoras y eficientes de procesamiento para la mejora del conocimiento y toma de decisiones en las organizaciones” (Aguilar & Álvarez, 2020).

Otra característica del análisis de *Big Data* es el uso de datos para una finalidad distinta de aquella para la que originalmente se habían recopilado; buena parte de reutilizar datos se basa en el concepto de *Open Data*, una herramienta gratis donde los datos están dispuestos para agregar valor y son de fácil accesibilidad (European Training Foundation, 2019). Durante 2012 y

2013 más del 60 % de los artículos de opinión de tecnología avanzada hablan de *Big Data* como la nueva estrategia indispensable para las empresas de cualquier sector, declarando poco menos, que aquellos que no se sumen a este nuevo movimiento se quedarán “obsoletas” en cuanto a la capacidad de reacción en sus decisiones, perdiendo competitividad y oportunidades de negocio contra su competencia.

2.2. Conceptos claves: Disciplinas y el *Big Data*

El *Big Data* por sí sólo no genera información predictiva cuantitativa, sin embargo, la unión de conocimientos de *Data Science* y *Data Analytics* si permite extraer valor del *Big Data* en forma de tendencias u oportunidades ocultas en la información. A partir de la llegada de la Inteligencia Operacional y *Business Intelligence* se ha descubierto una ventaja competitiva para aquellas empresas e instituciones que ya se dedican a aplicar la información y los datos a sus decisiones estratégicas (Nexus, 2022).

El uso de estas nuevas disciplinas está marcando la diferencia en diversos sectores, uno de ellos es el mercado bursátil, en donde las principales empresas de inversiones como JPMorgan o BlackRock, han establecido ya centros de investigación de *Big Data*, para estudiar la capacidad de esta herramienta y predecir el comportamiento del mercado sin técnicas tradicionales, tomando como fuente diferentes datos como indicadores macro o estados contables.

Figura 3. Disciplinas y el Big Data



Fuente: Elaboración propia a partir de DATAFLOQ, 2019.

Según la definición de Nexus (2022), Analítica de datos (*Data Analytics*) está relacionada con el uso de software para analizar los datos y donde se realizan procesos como: hallazgo de tendencias, patrones, correlaciones u otros

hallazgos útiles en el procesamiento del *Big Data*. Por lo tanto, la importancia de esta disciplina radica en traducir estos datos en información relevante y accionable para la toma de decisiones a partir de los datos y cómo esa información puede mejorar los procesos en análisis. A partir de esta disciplina, el *Big Data* genera valor a los procesos de un negocio o institución, así como también nuevas perspectivas que influirán en escenarios futuros a partir de tendencias actuales.

Asimismo, podría considerarse la Ciencia de datos (*Data Science*) como la ciencia que estudia los datos a través de modelos más eficaces que permiten la gestión y análisis de diferentes fuentes de datos. Ello implica un conjunto de disciplinas, que van desde la estadística, las matemáticas y la informática hasta la programación, pasando por procesos de creatividad que permiten mirar los posibles problemas desde perspectivas novedosas (Nexus, 2022). Dentro de los procesos de *Data Science* también aparecen disciplinas como *Machine Learning* o *Deep Learning*. Estos suponen la capacidad de generar sistemas cada vez más eficientes que “aprendan” y sean capaces de tomar decisiones en base a los datos.

2.2.1. Minería de datos / *Data Mining*

Según Kantardzic (2011), el *Data Mining* o minería de datos es un proceso de descubrimiento progresivo de conocimiento de una determinada colección de datos equilibrando el conocimiento de los expertos humanos en la descripción de problemas y objetivos con las capacidades de búsqueda de los equipos.

Por lo tanto, la minería de datos tiene dos objetivos, la predicción y la descripción. En el primero produce nueva información y conocimiento no trivial, mientras que el segundo produce modelos y roles de los sistemas. Dichos objetivos, se logran a partir del uso de técnicas de minería de datos y dentro de las más comunes se encuentran:

- ▶ **Clasificación:** Descubrimiento de una función de aprendizaje predictivo que clasifica un dato en una de varias clases predefinidas.
- ▶ **Regresión:** Descubrimiento de una función de aprendizaje predictivo que asigna un elemento de datos a una variable de predicción de valor real.

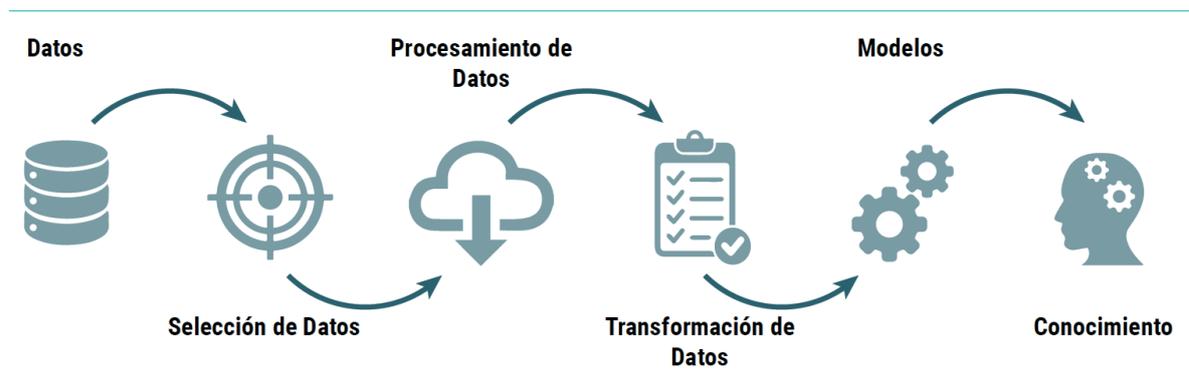
- ▶ **Agrupamiento:** Una tarea descriptiva común en la que se busca identificar un conjunto de categorías o clústeres para describir los datos.
- ▶ **Resumen:** Una tarea descriptiva adicional que implica métodos con una descripción compacta para un conjunto (o subconjunto) de datos.
- ▶ **Modelado de dependencias:** Búsqueda de un modelo local que describa dependencias significativas entre variables o entre los valores de una entidad en un conjunto de datos o en una parte de un conjunto de datos.
- ▶ **Detección de cambios y desviaciones:** Descubrimiento de los cambios más significativos en el conjunto de datos.

El procedimiento general adaptado a la minería de datos implica los siguientes pasos:

- ▶ **Establecer el problema o necesidad a partir de datos relevantes:** Este paso implica el conocimiento y la experiencia específica del problema, suelen ser necesarios para llegar a una declaración de problema significativa. Para ello se suele especificar un conjunto de variables que puedan explicar el problema y es necesario obtener datos de dichas variables.
- ▶ **Seleccionar datos:** Este paso se refiere a cómo se generan y recopilan los datos. En general hay dos posibilidades distintas: el primero es cuando el proceso de generación de datos está bajo el control de un experto, es decir, conlleva un diseño experimental y la segunda posibilidad, es cuando el experto no puede influir en los datos, por tanto, tiene un enfoque observacional, este paso es de gran importancia en el proceso, ya que puede afectar la elección del modelo y más adelante la interpretación final de los resultados.
- ▶ **Procesamiento de datos:** Incluye al menos dos tareas comunes, la detección y la eliminación de valores atípicos, los cuales pueden ser descritos como valores de datos inusuales que no son consistentes con la mayoría de las observaciones, los cuales pueden resultar de errores de medición, codificación y errores de registro y, a veces son naturales y con valores anormales.

- ▶ **Estimación del modelo:** La selección e implementación de la técnica de minería de datos adecuada es la tarea principal en esta fase. Este proceso no es sencillo ya que, por lo general, la implementación se basa en varios modelos, seleccionando el mejor según el conocimiento previo y los datos disponibles.
- ▶ **Interpretar el modelo y dar conclusiones:** En este proceso final se define nuevo conocimiento o bien, se afirma o niega una hipótesis previa frente al problema. En la mayoría de los casos, los modelos de minería de datos deberían ayudar en la toma de decisiones, por cuanto los modelos deben ser interpretables para ser útiles y es necesario tener en cuenta los objetivos del modelo y la precisión de su interpretación.

Figura 4. Proceso general de la minería de datos



Fuente: Elaboración propia a partir de Kantardzic, 2011.

La minería de datos descansa en las bases de la estadística, así como el aprendizaje automático, disciplinas que comparten el descubrimiento de conocimiento a partir de *Big Data* y aprendizaje de datos (Kantardzic, 2011).

2.2.2. Aprendizaje automático / *Machine Learning*

El aprendizaje automático o *Machine Learning* (ML) tiene como objetivo principal construir sistemas que puedan realizar o superar la competencia a nivel humano para aprender y mejorar la previsibilidad y el rendimiento automáticamente a través de la experiencia y los datos, sin ser programado para hacerlo por humanos.

Figura 5. Algunos usos actuales del *Machine Learning*



Fuente: Elaboración propia a partir de PROMIDAT, 2021.

El aprendizaje automático aborda el reconocimiento de patrones que pueda realizar o superar la competencia a nivel humano en el manejo de muchas tareas, problemas complejos e incluso, actividades recreativas a partir del uso de algoritmos programados. Dentro de los usos más comunes en la actualidad del aprendizaje automático se encuentran los procesos de funcionamiento de créditos, UBER, Spotify y Gmail tal como se evidencia en la Figura 5.

El aprendizaje automático se puede separar en tres principales tipos: estructurado, no estructurado y aprendizaje profundo (*Deep Learning*). Para el aprendizaje estructurado se entrena con un conjunto de ejemplos en los que los resultados de salida son conocidos. Los modelos aprenden de esos resultados conocidos y realizan ajustes en sus parámetros interiores para adaptarse a los datos de entrada.

Para el aprendizaje no estructurado se relacionan datos sin etiquetar cuya estructura es desconocida, de manera que el objetivo será la extracción de información significativa, sin la referencia de variables de salida conocidas, y mediante la exploración de la estructura de dichos datos sin etiquetar.

Y finalmente, el aprendizaje profundo o *Deep Learning* cuenta con una estructura jerárquica de redes neuronales artificiales, que se construyen de una forma similar a la estructura neuronal del cerebro humano, la cual permite abordar el análisis de datos de forma no lineal (PROMIDAT, 2021).

2.2.3. Minería web (*Web Mining*)

El término Minería web o *Web Mining* hace referencia al uso de técnicas de minería de datos para automáticamente descubrir y extraer información de documentos y servicios web (Kantardzic, 2011). Dentro de las técnicas utilizadas para la extracción de información disponible en la web se encuentran *Web Scraping* y *Web Crawling*.

El *Web Scraping* se utiliza para extraer datos que se encuentran estructurados en los sitios web. En las plataformas seleccionadas se realiza el “raspado” web mediante el diseño de un programa específico para cada sitio y país. Debido a que esta técnica implica abrir y cerrar las páginas de cada portal, imitando el comportamiento de los visitantes habituales, es necesario seguir un proceso determinado para que el ejercicio de recolección de datos sea lo menos invasivo para cada portal web (Sonia & Emanuel, 2021).

El *Web Crawling* se utiliza en sitios que no tienen disponible la información de forma estructurada. Las técnicas automatizadas de barrido en la web pueden ser programadas por programas como Python, utilizando las bibliotecas Selenium y BeautifulSoup. En el caso que no sea posible aplicar ninguna de las dos técnicas se procede a recopilar los datos manualmente, tal como puede ocurrir con algunos sitios web (Sonia & Emanuel, 2021).

2.2.3.1. *Web Scraping*

Consiste en la extracción de datos significativos de una o varias páginas web determinadas para una manipulación o análisis posterior, de forma que existen varias herramientas que permiten extraer datos web. Así pues, el *Web Scraping* permite:

- ▶ Alimentar una base datos
- ▶ Hacer una migración de un sitio web
- ▶ Recopilar y ofrecer datos dispersos por varias web
- ▶ Generar alertas

- ▶ Monitorear precios de la competencia
- ▶ Recolección de fichas de productos
- ▶ Detección de cambios en sitios web
- ▶ Analizar enlaces de sitios para buscar links

Existe cierta tensión entre el *Scraping* y los sitios web; cuanto más interesantes son los datos disponibles en un sitio Web, más protegidos tienden a estar. Por tanto, es conveniente revisar los términos legales de la página web antes de ejecutar estas acciones.

Las herramientas de *Web Scraping* están diseñadas para extraer información de sitios web de forma automática, son también conocidos como "*scrapeadores*". Estas herramientas resultan muy útiles para cualquiera que trate de recolectar datos de una página web, sin embargo, los usos más comunes y prácticos para los que se pueden utilizar son los siguientes:

- ▶ Extraer datos de contacto como por ejemplo email
- ▶ Extraer los títulos y contenidos de un blog
- ▶ Crear un canal RSS de los contenidos de una página web
- ▶ Seguir la evolución de precios de distintos productos

La principal ventaja de estas herramientas, presentadas a continuación, es que son sencillas de usar y no requieren grandes.

3 ▶ Usando el *Big Data* para procesar las Ofertas de Trabajo en Línea

Las Ofertas de Trabajo en Línea (OTL) son una rica fuente de información sobre habilidades y otros requisitos laborales que son difíciles de recopilar a través de métodos tradicionales. El acceso a esta información puede ayudar a los actores del mercado laboral a:

- ▶ Comprender mejor la demanda de habilidades y su dinámica.
- ▶ Permitir que las personas tomen mejores decisiones de carrera y desarrollo de habilidades.
- ▶ Apoyar a los empleadores en el desarrollo o ajuste de políticas de recursos humanos (RRHH).
- ▶ Ayudar a los formuladores de políticas a tomar decisiones más informadas.
- ▶ Mejorar la orientación de los servicios de empleo, y las instituciones de formación.

El análisis OTL puede proporcionar información adicional, detallada y oportuna sobre las tendencias del mercado laboral y permite la identificación temprana de trabajos y habilidades nuevas y emergentes. Sin embargo, el análisis OTL no reemplaza otros tipos de información e inteligencia del mercado laboral, de hecho, es más útil y representativo cuando se combina con otras fuentes convencionales como las encuestas de hogares u otras encuestas periódicas en el mercado de trabajo.

Un análisis significativo requiere considerar los siguientes desafíos:

- ▶ **Representatividad:** Las vacantes en algunos sectores y ocupaciones están sobrerrepresentadas en los portales de OTL, mientras que las de otros sectores están subrepresentadas. En estos casos las comparaciones con fuentes como las encuestas de hogares suelen ser reveladoras.

- ▶ **Integridad:** Las habilidades enumeradas en un aviso de vacante reflejan parcialmente el perfil completo del puesto: los empleadores tienden a enumerar solo las habilidades y calificaciones críticas para "atraer y filtrar" a los solicitantes de empleo.
- ▶ **Madurez:** Los patrones de uso de los portales de OTL difieren tanto entre países como dentro de ellos, lo que refleja el estado de la "brecha digital", la escasez de habilidades y estructuras de empleo particulares.
- ▶ **Simplificación:** Para ser utilizables en análisis de habilidades y del mercado laboral, los avisos de vacantes deben ser legibles por máquina y utilizar un vocabulario estandarizado y, dada la cantidad de datos, se deben hacer suposiciones simplificadoras.
- ▶ **Duplicación y estado de las vacantes:** El mismo aviso de vacante puede publicarse en varios sitios web y no necesariamente corresponde a una vacante real.

Tabla 3. Ventajas y desventajas del análisis de OTL

Ventajas	Desventajas
<ul style="list-style-type: none"> ▶ (Casi) Instantáneo o "en tiempo real" ▶ Gran volumen de información ▶ Tiempo y rentabilidad ▶ Información detallada sobre habilidades y necesidades de habilidades en todas las unidades y dentro de ellas (por ejemplo, países, ocupaciones) ▶ No es necesario recopilar "nuevos" datos ▶ La información en línea declarada por individuos puede ser más "veraz" 	<ul style="list-style-type: none"> ▶ La información no está estructurada y es imperfecta. Puede no tener representatividad estadística ▶ Errores de medición (por ejemplo, duplicación o prolongación de la vida útil de las vacantes) ▶ Preocupaciones de privacidad, consideraciones éticas/legales. ▶ Necesidad de habilidades analíticas de datos avanzadas (por ejemplo, software, programación, codificación) ▶ Análisis dificultado por la taxonomía seleccionada o los métodos de agrupación. ▶ Perfiles "parciales" de competencias profesionales.

Fuente: ILO, 2020

La publicación de vacantes en línea continúa creciendo, aunque no todos los tipos de ofertas de trabajo se publican en línea. Por ejemplo, se puede contratar a especialistas y ejecutivos; para algunos trabajos (por ejemplo, camareros), el empleador puede simplemente exhibir una nota en la ventana del lugar de trabajo, mientras que, para otros, los empleadores prefieren la contratación interna o el boca a boca.

Por lo tanto, es necesario intercambiar información, experiencias y buenas prácticas en la forma de una “Red de Expertos” que permitan desarrollar mejores estudios de *Big data* para el mercado laboral. Este intercambio debería contener al menos cuatro características; objetivos que soporten la planeación e implementación de *Big Data* para el análisis del mercado de trabajo, un plan de trabajo, modalidades de trabajo y de análisis usando el *Big data* que incluye las herramientas disponibles para recopilación de información y finalmente, la composición de equipos e iniciativas que vinculen a las partes interesadas que desarrollan *Big Data* para el análisis del mercado laboral.

Figura 6. Principales características de una red de expertos para desarrollar *Big Data* para IML (Información del Mercado Laboral)



Fuente: ILO, 2020

Las fuentes de vacantes varían en calidad y contenido. Para desarrollar una base de datos adecuada para el análisis posterior, el preprocesamiento implica las siguientes acciones:

- ▶ **Limpieza:** Los OTL están diseñados para atraer al candidato más adecuado, no para proporcionar datos limpios para el análisis del mercado laboral. Además de información útil desde el punto de vista analítico, el análisis OTL también puede descargar "ruido" (como anuncios, opciones no marcadas en menús desplegados, presentaciones de perfiles de empresas, etc.). La limpieza implica una secuencia de actividades para eliminar el "ruido" de los datos para mejorar su calidad y prepararlos para las siguientes fases.
- ▶ **Fusión:** Los empleadores a menudo publican una vacante en más de un portal. Las personas que agregan (páginas web que recopilan vacantes de varios portales de empleo) aumentan la posibilidad de encontrar duplicados. Las duplicaciones de vacantes no son deseables para el análisis final, pero en la parte inicial de la fase de procesamiento previo pueden enriquecer los datos al permitir recopilar y combinar información diferente sobre la misma vacante publicada en diferentes lugares.
- ▶ **Eliminación de duplicados:** Una vez que se han fusionado los datos de la misma vacante, es necesario eliminar las vacantes duplicadas del análisis. Una OTL se considera un duplicado si la descripción y la ubicación del trabajo son las mismas que las de otro anuncio de trabajo en la base de datos. Los metadatos de vacantes (como ID de referencia, URL de página) también se pueden usar para identificar y eliminar duplicados de vacantes en sitios web de agregadores.

La información de las vacantes se extrae utilizando algoritmos de aprendizaje automático y basados en ontologías⁴ que se basan en un marco preexistente para el procesamiento y análisis de vacantes en línea.

4 Se refiere a un modelo capaz de representar, mediante el uso de un lenguaje *machine-learning*, formal y basado en la lógica matemática, los aspectos relevantes de un fenómeno o de un ámbito de interés.

Por ejemplo, el sistema del Centro Europeo para el Desarrollo de la Formación Profesional (CEDEFOP), utiliza ontologías estándar: la ESCO (Clasificación Europea de Ocupaciones, habilidades y competencias), CEAE (Clasificación Estadística de Actividades Económicas) para industria, NUTE (Nomenclatura de Unidades Territoriales de Estadística) para lugares de trabajo, CINE (Clasificación Internacional Normalizada de la Educación) para nivel educativo y, ontologías personalizadas (por tipo de contrato, experiencia, salario, horas de trabajo) relacionadas con habilidades y el mercado laboral.

Las ontologías se actualizan y enriquecen continuamente, sobre la base de procesos automáticos y validación experta y en los casos en que no se obtienen los resultados, la información se extrae utilizando un algoritmo de aprendizaje automático, entrenado sobre la base de grandes conjuntos de datos para ajustar mejor las variables en un idioma determinado. En el sistema del CEDEFOP, actualmente se utiliza el aprendizaje automático para clasificar solo las ocupaciones; se está desarrollando la aplicación del aprendizaje automático a otras variables.

La tabla 4, muestra algunas de las principales habilidades técnicas, blandas y digitales identificadas en pronósticos de este tipo. Es importante aclarar que las listas cambiarán según el contexto (industria o actividad) en el que se aplican.

Tabla 4. Habilidades más demandadas resultado de *webscrapig* para Paraguay 2021

Técnicas	Digitales	Blandas
Gestión de proyectos	Micro Office	Proactividad
Control de calidad	Programación	Responsabilidad
Marketing y Ventas	Teletrabajo	Liderazgo
Contabilidad	Tic	Trabajo en Equipo
Inglés	Robótica	Comunicación

Fuente: Tomado de los resultados del Dashboard de Paraguay <https://n9.cl/paraguay> <https://www.oitcinterfor.org/cambios-dinamicaslaborales-paraguay-2020-2021>.

3.1. Recomendación para una mejor identificación de habilidades en los ejercicios *Big data*

Un reciente estudio de OIT encontró que las OTL son una fuente promisoría para el análisis de la dinámica de los perfiles y habilidades. Los datos por país pueden arrojar nuevas perspectivas para abordar la configuración de los perfiles demandados en el mercado laboral. No obstante, la representatividad de estos datos no está asegurada y es muy recomendable complementar su análisis con otras fuentes o en segmentos específicos del mercado⁵.

Es necesario definir conjuntos de habilidades compatibles, significativas y útiles que se basen en un contexto semántico relevante, y no en palabras clave y cadenas, y que adicional dichas habilidades sean detalladas, diferenciando por grupos ocupacionales. Para ello, en muchos casos, los propietarios de los portales de Ofertas de Trabajo en Línea (OTL) otorgan acceso directo a los datos a través de IPA (Interfaz de Programación de Aplicaciones), pero cuando esto no es posible, se utilizan método como el raspado y el rastreo.

Un ejemplo práctico es el Skills HAVLE (Herramienta de Análisis de Vacantes en Línea para Europa) que utiliza tres métodos para recopilar OTL:

1. El primer y más importante método es el acceso directo a través de la **Interfaz de Programación de Aplicaciones (API)**, que permite descargar el contenido de las vacantes directamente desde las bases de datos del portal OTL. Este acceso directo requiere un acuerdo formal con el operador del sitio web e incurre en costos de mantenimiento y acuerdo; los datos recopilados de esta manera son de mayor calidad y se pueden descargar mucho más rápido que los datos recopilados por otros métodos.
2. **El raspado (*scraping*)** también se utiliza para extraer datos estructurados de sitios web y se basa en datos que ya están estructurados en la página web para que puedan extraerse con precisión al conocer la posición exacta de cada campo en la página. Como se deben programar ejercicios de "*scraping*" específicos para

5 Using Online Vacancy and Job Applicants' Data to Study Skills Dynamics. ILO. 2022.

cada sitio web, es un método ideal para sitios que contienen muchas vacantes.

- 3. El rastreo** se usa solo como último recurso ya que este método utiliza un robot programado para navegar sistemáticamente por los portales web y descargar sus páginas, es mucho más general que el *scraping* y es más fácil de desarrollar. Sin embargo, los rastreadores recopilan significativamente más ruido del sitio web (contenido irrelevante), por lo que se necesita más esfuerzo para limpiar los datos antes de continuar con el procesamiento.

3.2 Pasos claves para el procesamiento de Bases de Datos

Existen pasos claves en el proceso de Descubrimiento de Conocimiento en Bases de Datos (DCBD) a partir de la perspectiva técnica y la perspectiva del mercado laboral en donde se considera un equipo interdisciplinar las características más relevantes del "V" de *Big Data* a través de los siguientes pasos a partir de (ILO, 2020);

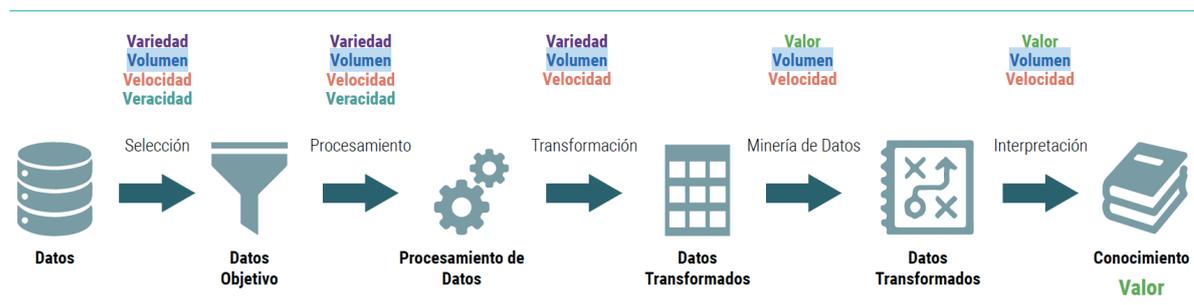
Paso 1. Selección de datos: Para plantear objetivos claros y en orden al procesamiento de datos por parte de un equipo de expertos en la fase de selección de datos incluye lo siguiente:

- ▶ Identificar los criterios que deben ser incluidos en el modelo de origen, así como sus fuentes y su clasificación (Expertos en estadística).
- ▶ Identificar un paradigma de modelo de datos apropiado para almacenar grandes volúmenes de datos y la forma en que estos se recopilan (Expertos técnicos).
- ▶ Reconocer el acceso a los datos y el proceso necesario para programar automáticamente la recopilación de datos.
- ▶ Identificar criterios para la selección y validación de las fuentes pertinentes (Expertos del mercado laboral).

Paso 2. Procesamiento de datos: Entre las principales acciones necesarias en esta segunda fase figuran las siguientes:

- ▶ Definir criterios para evaluar la consistencia, precisión e importancia de los datos (Expertos estadística).
- ▶ Establecer un proceso para la identificación de los registros de datos duplicados y valores faltantes (Expertos técnicos- Expertos mercado laboral).
- ▶ Identificar los sinónimos utilizados en el ámbito del mercado laboral y contribuir así a mejorar la precisión de los datos (Expertos mercado laboral).

Figura 7. El proceso DCBD y las grandes “V” de Big Data involucradas en cada paso



Fuente: Elaboración propia a partir de ILO, 2020.

Paso 3. Transformación de datos: Las preguntas clave planteadas por la fase de transformación son:

- ▶ Para los expertos en estadística: ¿Cómo medir la integridad del modelo objetivo identificado? ¿El modelo objetivo mantiene la importancia de los datos al final del proceso ETL?
- ▶ Para los expertos técnicos: ¿Cómo desarrollar procedimientos de Big Data para transformar los datos en bruto en un objetivo modelo de una manera escalable?
- ▶ Para los expertos en el dominio del mercado laboral: ¿Cómo identificar el formato y la taxonomía de los datos de destino?

Paso 4: Minería de datos (*Data Mining*) & *Machine Learning*: Las preguntas claves se orientan a:

- ▶ Para los expertos estadísticos y técnicos: ¿Cómo seleccionar los mejores algoritmos? ¿Cómo afinar sus parámetros? ¿Cómo evaluar la efectividad del algoritmo? ¿Cómo implementar a escala?
- ▶ Para los expertos en el ámbito del mercado laboral: ¿Qué conocimientos se deben seleccionar y cuáles se deben descartar? ¿Cuál es la importancia para el mercado laboral de los conocimientos obtenidos? ¿Qué ideas novedosas ha descubierto de esta manera? ¿Cómo explicar los resultados del proceso minero de un mercado laboral y su perspectiva?

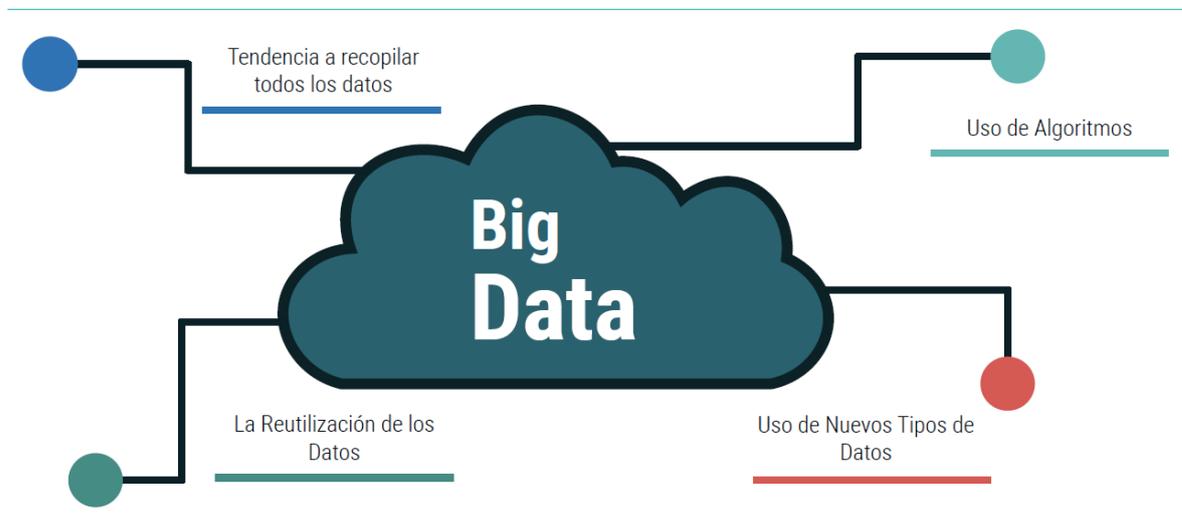
Paso 5. Evaluación e interpretación: Las preguntas clave planteadas por esta fase final incluyen:

- ▶ Para los expertos estadísticos y técnicos: ¿Cómo seleccionar el paradigma de visualización? ¿Cómo seleccionar un modelo de visualización apropiado para el conocimiento que queremos retratar?
- ▶ Para los expertos en el ámbito del mercado laboral: ¿Cómo entregar el conocimiento adecuado de acuerdo con las necesidades de las partes interesadas? ¿Cómo identificar las rutas de navegación visual para cada parte interesada? ¿Cómo recuperar comentarios (si los hubiera) de los usuarios del mercado laboral? ¿Cómo insertar el conocimiento del mercado laboral en el negocio del medio ambiente?

3.3 Parámetros del *Big Data*

El funcionamiento actual del *Big Data* está basado en parámetros estadísticos que hacen distintivo su análisis en temas como el uso de algoritmos, el uso de datos, la reutilización de datos y el uso de nuevos tipos de datos, tal como se muestra en la Figura 8.

Figura 8. Parámetros estadísticos de *Big Data*



Fuente: Elaboración propia a partir de ICO, 2018.

3.3.1. Uso de algoritmos

Tradicionalmente, el análisis de un conjunto de datos implica, en términos generales, decidir qué se quiere saber a partir de los datos y construir una consulta para encontrarla, identificando lo relevante. En el análisis de *Big Data*, por otro lado, generalmente no comienza con una consulta predefinida para probar una hipótesis particular, a menudo implica una “fase de descubrimiento” de ejecutar un gran número de algoritmos contra los datos para encontrar correlaciones (European Training Foundation, 2019).

Esta es una forma de aprendizaje automático, ya que el sistema “aprende” cuáles son los criterios pertinentes a partir del análisis de los datos. Si bien el uso de algoritmos no es un tema nuevo, su uso de esta manera es una característica de análisis de *Big Data*.

	ESTADÍSTICA TRADICIONAL		BIG DATA
USO DE ALGORITMOS	Pensar con datos Decidir lo que se quiere saber a partir de los datos recolectados.	VS.	Actuar con datos Ejecutar un gran número de algoritmos contra los datos para encontrar correlaciones.

3.3.2. Uso de datos

Para analizar los datos con métodos de estadística tradicional es necesario encontrar una muestra estadísticamente representativa o llevar a cabo muestreo aleatorio. Mientras que *Big Data*, implica un enfoque de capacidad computacional, que con modelos tradicionales estadísticos no se podrían operar por el volumen de datos. Esta característica de la analítica de *Big Data* ha sido más fácil gracias a la capacidad de almacenar y analizar cada vez más cantidades de datos.

USO DE DATOS	ESTADÍSTICA TRADICIONAL	VS.	BIG DATA
	Pensar con datos Decidir lo que se quiere saber a partir de los datos recolectados.		Actuar con datos Analizar todas las entradas de información, todos los datos disponibles.

3.3.3. Reutilización de datos y nuevos tipos de datos

Una característica primordial del análisis de *Big Data* es el uso de datos para una finalidad distinta de aquella para la que originalmente se habían recopilado, esto se debe a que el análisis es capaz de extraer datos para nuevos conocimientos, encontrar correlaciones diferentes a las supuestas y relacionar conceptos aparentemente dispares.

Por otra parte, desarrollos en tecnología como *Internet of Things*⁶ (IoT), junto con los desarrollos en el poder de la analítica de *Big Data*, significan que el escenario tradicional en el que las personas proporcionan conscientemente los datos personales ya no son la única o principal forma en que los datos personales se recopilan datos (European Training Foundation, 2019).

⁶ Internet de las cosas, describe la red de objetos que conectan e intercambian datos con otros dispositivos y sistemas a través de Internet.



Seguimiento en Línea



Búsquedas en Internet



Sensores en la Calle



Uso de Energía Eléctrica



Dispositivos de Monitoreo

En muchos casos, los datos que se utilizan para el análisis se han generado automáticamente, por ejemplo, mediante el seguimiento en línea, búsquedas en internet, sensores en la calle, uso de energía eléctrica y uso dispositivos de monitoreo.

3.3.4. Taxonomía de los datos

Los datos utilizados en el análisis de *Big Data* pueden recopilarse a través de nuevos canales, pero alternativamente pueden ser nuevos datos producidos por el análisis, en lugar de ser proporcionados conscientemente por individuos.

Por lo tanto, la tipificación de datos (ICO, 2018) pueden clasificarse en: proporcionados, observados, derivados e inferidos.

- ▶ Los **datos proporcionados** son datos conscientemente por individuos, por ejemplo, cuando relleno un formulario online.
- ▶ Los **datos observados** se registran automáticamente, por ejemplo, mediante cookies en línea o sensores o CCTV vinculados al reconocimiento facial.
- ▶ Los **datos derivados** se producen a partir de otros datos de una forma relativamente sencilla y la moda directa, por ejemplo, calcular la rentabilidad del cliente desde el número de visitas a una tienda y artículos comprados.
- ▶ Los **datos inferidos** se producen mediante el uso de un método más complejo de análisis para encontrar correlaciones entre los conjuntos de datos y el uso de estos para categorizar o perfilar a las personas, por ejemplo, calcular los puntajes de crédito o predecir los resultados de salud futuros. Los datos inferidos se basan en probabilidades y, por lo tanto, se puede decir que son menos “ciertos” que datos derivados.

3.4. Experiencias internacionales del Big Data en la identificación de brechas de habilidades y de talento humano

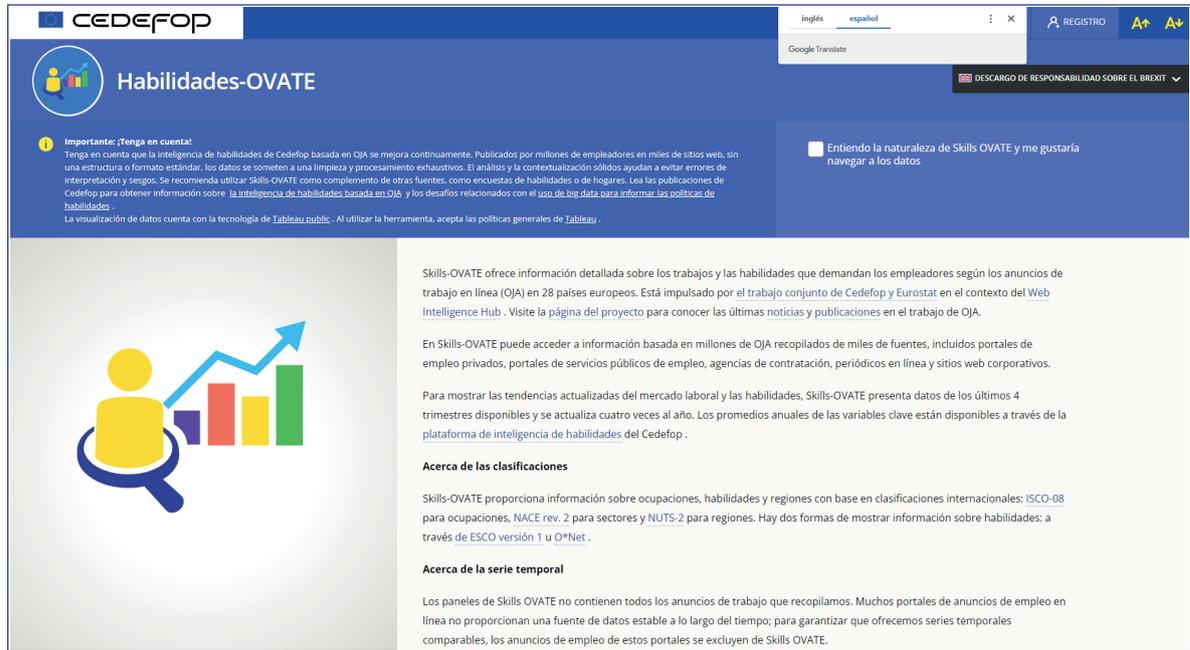
Esta sección muestra algunos de las iniciativas puestas en acción por algunos países y entidades para el análisis del mercado de trabajo y de OTL haciendo uso del *Big Data* (recopilados en el Anexo 1. **Resumen de estudios aplicativos de Big Data en el mercado laboral**).

3.4.1. Sector público

3.4.1.1. Análisis de vacantes

Un primer ejemplo es el del Centro Europeo para el Desarrollo de la Formación Profesional -CEDEFOP- en su herramienta en línea Skills-OVATE *Online Vacancy Analysis Tool for Europe*, reúne mediante Web Scrapping/ Crawling datos de portales de empleo privados, portales de servicios públicos de empleo, agencias de reclutamiento, avisos de periódicos en línea y páginas de empleadores. Su objetivo es estimar el grado de representatividad de las vacantes de empleo en línea, para ello, mapean y analizan sitios de vacantes locales de empleo en línea en 28 países miembros de la comunidad europea. La información no se presenta como un índice con la evolución de las vacantes sino como una síntesis de datos de millones de vacantes de empleo de línea. En general, para el 2021 las vacantes representan entre 50 % (Dinamarca, Grecia y Rumania) a 100 % (Estonia, Finlandia y Suecia) del total de anuncios en el país, aunque posiblemente estas proporciones se incrementen conforme a las actualizaciones de vacantes.

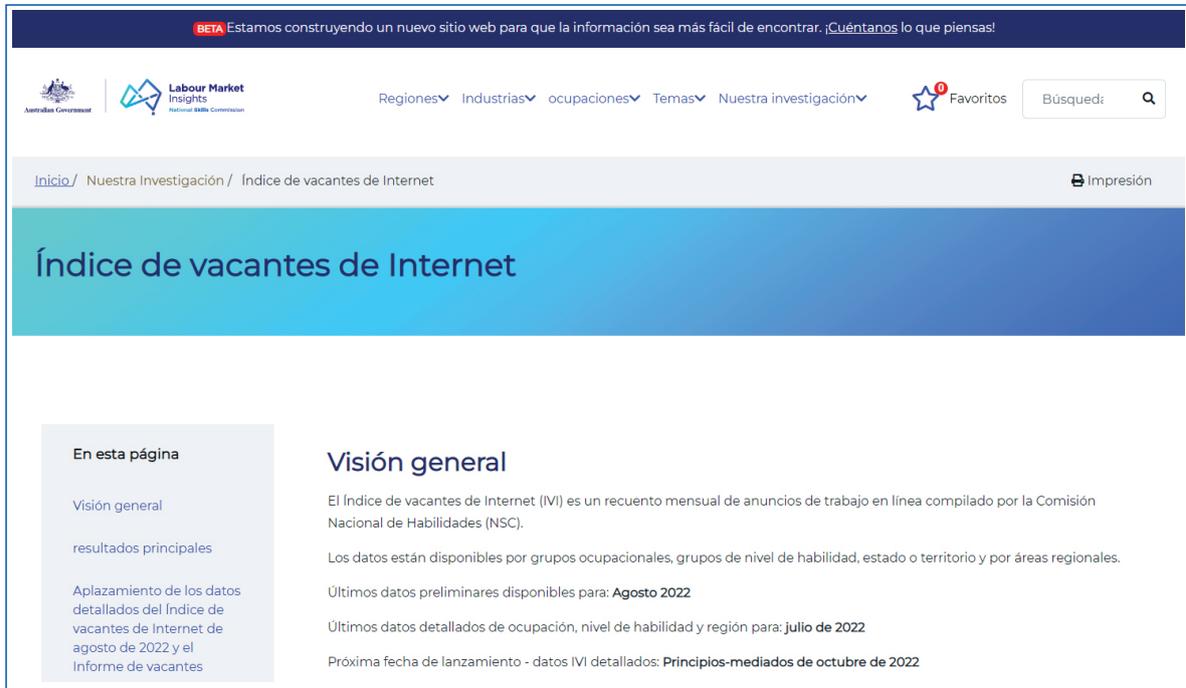
Figura 9. Sitio web Skills-OVATE



Fuente: Skills-OVATE | CEDEFOP (europa.eu).

Por otra parte, el Ministerio de Trabajo y Pequeñas Empresas de Australia en 2021, estimó el Índice de vacantes de Internet (Index Internet Vacancy), con la finalidad de monitorear la evolución de vacantes de empleo en su territorio. Este indicador es la única fuente disponible pública de datos detallados sobre vacantes en línea. Se basa en un recuento de anuncios de trabajo en línea presentados mensualmente en SEEK, CareerOne y Australian JobSearch. El gobierno de Australia presenta datos mensuales del comportamiento de las vacantes a través de mapas de calor, dashboard, tablas Excel y gráficas. Para enero de 2022 se presentó un crecimiento actual de vacantes del 40,8 % en comparación a enero 2021.

Figura 10. Sitio web Internet Vacancy Index



Fuente: Internet Vacancy Index | Labour Market Insights.

3.4.1.2. Análisis de competencias

La experiencia del Ministerio de Empresa, Innovación y Empleo de Nueva Zelanda, usa la herramienta "*Job vacancies*" que permite concatenar opciones de vacantes con carreras y habilidades, adicionalmente, hace estudio del comportamiento de estas variables a lo largo del tiempo. Por otro lado, la oferta de trabajo se estima a través de *Jobs Online* mediante el uso de anuncios de trabajo como un proxy para las vacantes de trabajo en línea de cuatro bolsas de trabajo, *Jobs Online* proporciona un indicador clave de la demanda de mano de obra, vacantes por industria, ocupación, nivel de habilidades y región. En este sentido, la meta del ministerio es generar una herramienta de educación en línea y exploración de carreras en perspectiva de la ocupación, dada la evaluación de la oferta y la demanda de empleo.

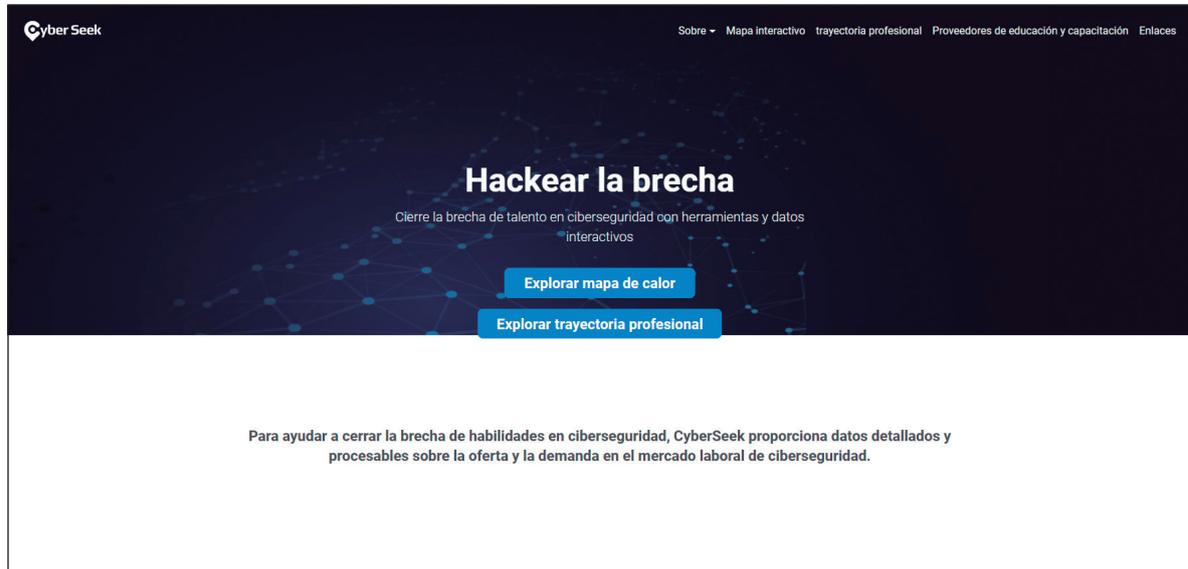
Figura 11. Sitio web Current Board Vacancies



Fuente: Current board vacancies | Ministry of Business, Innovation & Employment (mbie.govt.nz).

El análisis de empleo en el sector de ciberseguridad de Estados Unidos observó cientos de millones de ofertas de trabajo y transiciones profesionales de la vida real para proporcionar información sobre los patrones del mercado laboral. Con el objetivo identificar habilidades y brechas; combinó mediante mapas de calor habilidades y herramientas, ocupaciones, títulos de trabajo y certificaciones que se encuentran en las ofertas de trabajo para definir un panorama más amplio de la ciberseguridad. Las vacantes guardaban relación con ocupaciones y certificaciones específicas de cada sector, acorde a los datos suministrados por el gobierno para los años 2017 a 2021. El análisis encontró que, entre octubre de 2020 hasta septiembre de 2021, hubo 162,700 vacantes para analistas de seguridad de la información, pero solo 138,000 trabajadores actualmente empleados en esos puestos. Hay 435,067 vacantes adicionales que solicitan habilidades relacionadas con la ciberseguridad, y los empleadores están luchando por encontrar trabajadores que las posean. En general, los roles de ciberseguridad tardan en promedio un 21 % más en llenarse que otros trabajos de TI.

Figura 12. Sitio web Cyber Seek



Fuente: Cyberseek.

3.4.2. Sector privado

3.4.2.1. Demanda Laboral

En este escenario, la universidad de Oxford con el iLabor Project, estimó el índice de empleo en línea, para el cual recopilamos información sobre vacantes de empleos globales de las plataformas de trabajo en línea por país y ocupación. Es el primer indicador económico que proporciona una economía de trabajo en línea equivalente a las estadísticas del mercado laboral convencional y permite medir la oferta y la demanda de mano de obra independiente en línea en todos los países y ocupaciones mediante el seguimiento de la cantidad de proyectos y tareas en todas las plataformas en tiempo real. Los resultados son publicados a través de bases de datos, mapas de calor, dashboards de fácil acceso. Para 2022 Estados Unidos tenía como top de vacantes en desarrollo de software y creativos multimedia, mientras que en Colombia las más solicitadas fueron creativos en multimedia, desarrollador de software y traductores.

Figura 13. Sitio web The iLabourProject

El Proyecto iLabour
 Investigando la Construcción de Mercados Laborales, Instituciones y Movimientos en Internet

OXFORD INTERNET INSTITUTE | UNIVERSITY OF OXFORD

HOGAR | **ÍNDICE LABORAL EN LÍNEA** | INVESTIGAR | PUBLICACIONES | PRENSA | GRUPO DE PROYECTO | BLOG

El índice laboral en línea

El índice laboral en línea (OLI) es el primer indicador económico que proporciona una economía de trabajo en línea equivalente a las estadísticas del mercado laboral convencional. Mide la oferta y la demanda de mano de obra independiente en línea en todos los países y ocupaciones mediante el seguimiento de la cantidad de proyectos y tareas en todas las plataformas en tiempo real.

Una nueva versión del Índice Laboral en Línea está disponible en www.onlinelabourobservatory.org. La versión anterior disponible a continuación ya no se mantiene activamente.

OLI | By occupation | By employer country | Occupation x empl country

Búsqueda

El índice laboral en línea

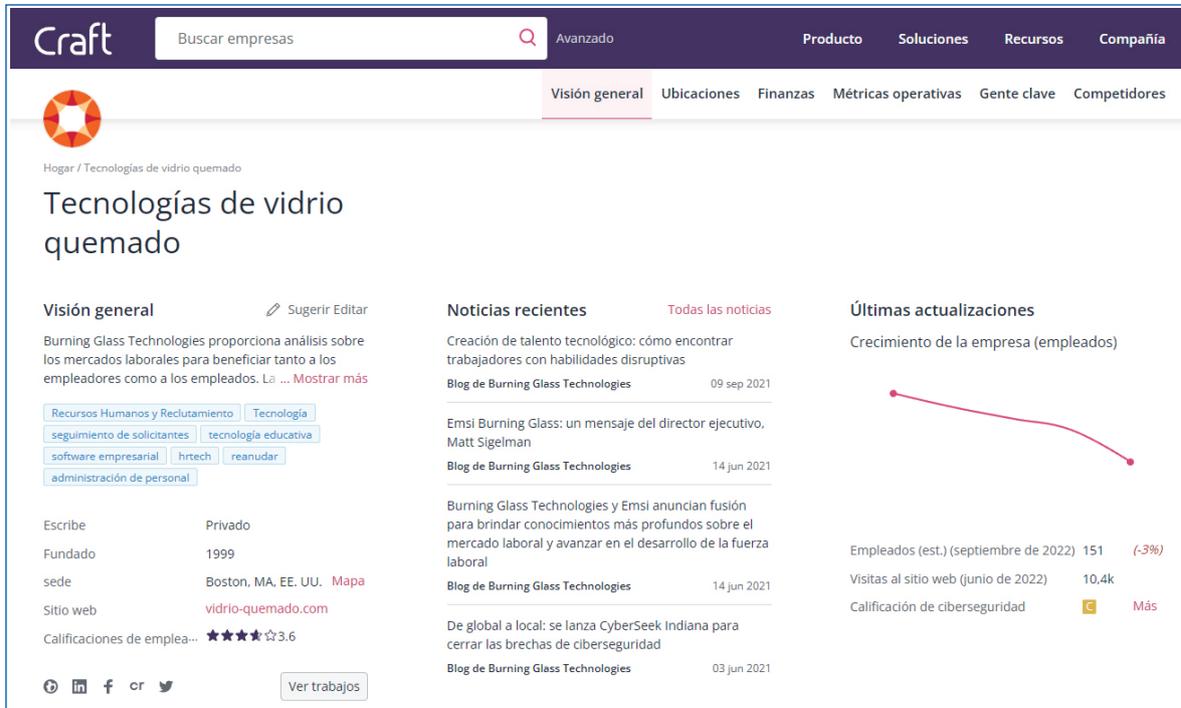
Gráfico de líneas que muestra el Índice Laboral en Línea (OLI) desde el 15 de mayo hasta el 22 de agosto. El eje Y representa el Índice Laboral en Línea, con marcas en 95, 100, 105 y 110. La línea muestra una tendencia general ascendente, comenzando cerca de 100 en mayo y alcanzando aproximadamente 110 en agosto.

El índice laboral en línea (OLI) es el primer indicador económico que proporciona una economía de trabajo en línea equivalente a las estadísticas del mercado laboral convencional.

Fuente: The Online Labour Index | The iLabour Project (ox.ac.uk).

En Estados Unidos, “*Burning Glass Technologies*” (servicio de pago) provee información de vacantes, que está detallada por sector, ocupación, y a nivel de regiones o ciudades. La información recopilada es de miles de plataformas de empleo en el país y las complementa con otras fuentes de vacantes de empleo, esta plataforma proporciona datos detallados y localizados del mercado laboral lo cual permite a los educadores aumentar la inscripción y fomentar el éxito de los estudiantes, mientras que las agencias de fuerza laboral aumentan la visibilidad de las necesidades del mercado laboral local y los departamentos de recursos humanos mejoran las decisiones relacionadas con el talento. En general, se han empleado para instituciones educativas como La Universidad de Virginia, donde se utilizaron datos del mercado laboral para el diseño y mercadeo de programas. Asimismo, gobiernos como el del Estado de Carolina del Norte quienes utilizaron los datos del mercado laboral para validar y reestructurar los programas de educación superior.

Figura 14. Sitio web Burning Glass Technologies

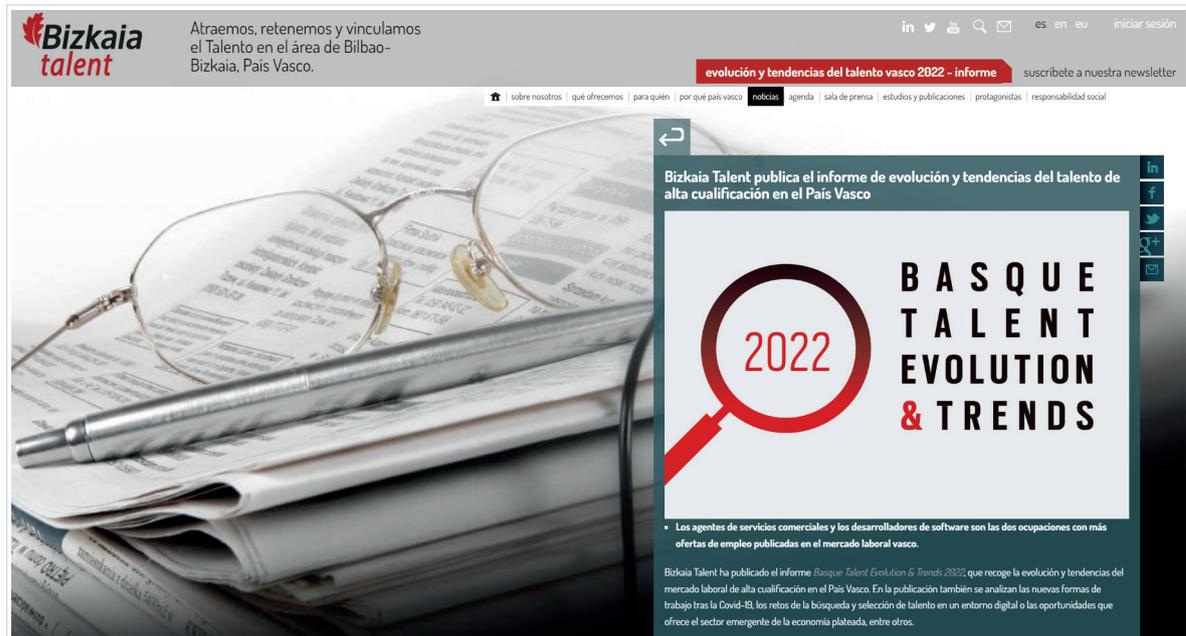


Fuente: Burning Glass Technologies Company Profile - Office Locations, Competitors, Revenue, Financials, Employees, Key People, Subsidiaries | Craft.co.

El Observatorio de talento Bizkaia Basque de España, relaciona la oferta educativa acorde a los perfiles que requiere el mercado. Esta es una iniciativa llevada a cabo por Bizkaia Talent junto con TabulaeX, spin-off de la Universidad de Milano-Bicocca, creada para gestionar su transferencia de conocimiento, basada en una herramienta que analiza el mercado laboral vasco centrado en profesionales de alta cualificación a través del análisis de *Big Data* de múltiples fuentes online; para ello, se emplean los portales de empleo, universidades y datos del gobierno.

Cuenta con 34.030 ofertas de empleo desde el 1 de enero de 2017 al 31 de diciembre de 2018, con lo que Bizkaia se convierte en el territorio con más ofertas de alta cualificación en cada uno de los dos años y con una tendencia al alza, ya que ha pasado de contar un 49 % en 2017 a un 54 % en 2018. De todas ellas, es el sector servicios el de mayor oferta pública (75,3 % en 2017 y 72,3 % en 2018), muy por delante del sector de la industria, el cual registró un gran crecimiento de publicación de ofertas pasando del 16,3 % en 2017 al 21,1 % en 2018.

Figura 15. Sitio web Bizkaia Talent



Sitio web: Bizkaia Talent publica el informe de evolución y tendencias del talento de alta cualificación en el País Vasco.

3.4.2.2. Análisis de brechas

La Universidad Tecnológica de Sydney (UTS) y la UNSW Sydney han desarrollado un método basado en el aprendizaje automático que puede identificar y recomendar trabajos con conjuntos de habilidades subyacentes similares a los actuales de alguien. El sistema también puede responder en tiempo real a los cambios en la demanda laboral y brindar recomendaciones sobre las habilidades precisas necesarias para la transición a una nueva ocupación.

Con el objetivo de desarrollar un enfoque de aprendizaje automático capaz de predecir la escasez de mano de obra para las ocupaciones, mediante el predictivo *Machine Learning*, se lograron analizar los conjuntos de habilidades subyacentes de más de 8 millones de empleos anunciados en Australia entre 2012 y 2020. También se compararon las predicciones de la transición laboral con los datos de la encuesta de Dinámica del hogar, los ingresos y el trabajo en Australia (HILDA), que realiza un seguimiento de los participantes a lo largo de sus vidas, para validar estas predicciones con casi 3000 ejemplos de la vida real.

3.4.3. Casos internacionales

La generación de herramientas para el análisis de habilidades es una tendencia mundial. Como ejemplo de eso, la UNESCO⁷ mediante el proyecto <https://my.job.com/> en Malawi, recopila información sobre las vacantes más demandadas y sus características de plataformas de trabajo en línea por país y ocupación. Es el primer indicador económico que proporciona una economía de trabajo en línea equivalente a las estadísticas del mercado laboral convencional. Mide la oferta y la demanda de mano de obra independiente en línea en todos los países y ocupaciones mediante el seguimiento de la cantidad de proyectos y tareas en todas las plataformas en tiempo real. Los resultados son publicados a través de bases de datos, mapas de calor, dashboards de fácil acceso, para 2022 Estados Unidos tenía como top de vacantes en desarrollo de software y creativos multimedia, en comparación con Colombia las más solicitadas fueron creativos en multimedia, desarrollador de software y traductores.

En cuanto a la intermediación laboral, la OIT en el Primer informe: desajustes entre demanda y oferta de trabajo, expone una mirada desde las ocupaciones. y se ejecuta como un monitor del mercado laboral para 2021, con base en datos de plataformas como Buscojobs y portales web de búsqueda de empleo. Para esto, empleó la técnica de *Web Scraping* con el objetivo de aportar información de calidad para el debate y diseño de política pública y abordar problemas estructurales del mercado de trabajo. El estudio concluyó que los portales de búsqueda de empleo son fuente de información valiosa, dado que, el mercado laboral no es un todo homogéneo y la lectura por ocupaciones aporta valor para diferenciar instrumentos de política.

Se encontraron además indicios de que en un número importante de ocupaciones existe descalces de habilidades y potencialmente escasez: menor inversión y bajos niveles de productividad. El informe también proporciona orientaciones para up-skilling y re-skilling en población de mayores de 45 años y es una guía para servicios públicos de orientación educativa y laboral e intermediación laboral.

7 <https://www.stepmalawi.com/>

Figura 16. Sitio web Monitor Laboral



Fuente: <https://ced.uy/>- Monitor Laboral: Desajustes entre oferta y demanda en el mercado laboral uruguayo: una mirada desde las ocupaciones. Uruguay: Análisis de oferta y demanda de empleo a partir de bases de datos a 4 meses de la pandemia COVID-19 | OIT/Cinterfor (oitcinterfor.org).

Para el **análisis de brechas de habilidades**, OIT/Cinterfor llevó a cabo un estudio en la República Dominicana en 2021 en conjunto con el Ministerio del Trabajo. Dicho estudio se basó en una propuesta metodológica que dimensiona la brecha de habilidades a partir de desajustes percibidos por el comportamiento de las ocupaciones en la interacción oferta-demanda. Se identificaron vacantes en ocupaciones de alta rotación, vacantes de difícil consecución (dificultad para ser cubiertas) y vacantes neutras o de comportamiento normal. Como resultado se obtuvo que cerca de la mitad de las vacantes requieren un nivel de cualificación alto (45,5 %), el 34,2 % medio y el 18,8 % bajo. En este caso, se detectó que hasta el 60 % de vacantes requieren hasta dos años de experiencia y el 30 % de las vacantes requieren solo un mes como experiencia mínima. Las palabras o frases más comunes en la descripción de las vacantes son “trabajo en equipo”, “construcción de relaciones”, “liderazgo”, “gestión de personas”, “organización” y “comunicación eficiente”.

Figura 17. Estudio de análisis de oferta y demanda de empleo, República Dominicana



Fuente: República Dominicana: análisis de oferta y demanda de empleo y trazabilidad de egresados Infotep, a partir de bases de datos | OIT/Cinterfor (oitcinterfor.org).

En cuanto a la **relación oferta y demanda de laboral**, el proyecto CEPAL de seguimiento de vacantes de empleo en línea en América Latina y el Caribe, generó un índice de vacantes, para evaluar el comportamiento de la oferta laboral sectorial. Los datos considerados contemplan los portales seleccionados: Bumeran, CaribbeanJobs, CaribbeanJobsOnline, CompuTrabajo, Freelancer, JobisJob, Profdir y Upwork.

Este proyecto es soportado con la a técnica *Scrapping-Crawling*, en el que se realiza recolección y caracterización de vacantes de empleo en línea.

En la primera etapa se seleccionaron las fuentes de información.

En segundo lugar, se analizaron las fuentes que resultaron seleccionadas para determinar cómo se organizaba la información de modo de poder aplicar las técnicas que permitan descargar los datos en línea (Scraping y crawling).

Posteriormente se realizó el procesamiento de los datos recopilados lo que implicó desarrollar una serie de tareas de limpieza y estructuración de estos.

Por último, se seleccionó algunas herramientas que mejor se ajustan para la visualización y análisis de los datos. Se concluyó que, a lo largo del año 2020, se observó una reducción en el número de vacantes de empleo globales durante los primeros meses de la pandemia en las tres categorías más importantes tecnología, marketing y ventas.

Durante el 2020 y hasta mediados de 2021, la mayoría de las vacantes globales de empleo del sitio freelancer y upwork estaban concentradas en dos sectores: servicios de TIC, diseño de páginas web, desarrollo de software y diseño, comunicación, ingeniería y ciencia de datos. En el caso de freelancer le siguen los servicios redacción, traducción e idioma y en el caso de upwork los servicios de venta y marketing.

▶ Conclusiones

El estudio de las ofertas en línea se ha expandido rápidamente en los últimos años, aunque se sigue centrando en países individuales, con un pequeño número de intentos de trabajo comparativo de nivel regional. La mayoría de las investigaciones se centran en comprender la demanda laboral, a partir del análisis de las vacantes en línea, esto obedece generalmente a la facilidad en el acceso a la información sobre puestos vacantes en comparación a la información de los candidatos, que algunos portales ofrecen de forma paga.

Este campo de investigación continúa siendo impulsado por un enfoque en las economías avanzadas, principalmente Estado Unidos y la Unión Europea; pero, también existe una tendencia evidente hacia la expansión de esta investigación a los mercados laborales de los países de renta media. Estos estudios aún siguen presentados desafíos debido a los sesgos que surgen principalmente por los diferentes niveles de acceso a Internet y desarrollo de la economía digital; también por las formas y modalidades utilizadas para anunciar una vacante, entre diferentes sectores o empresas de distintos tamaños; así como mayores niveles de informalidad en algunas economías o sectores.

Las Ofertas de Trabajo en Línea (OTL) abren nuevas vías para comprender la evolución de los mercados laborales. La disponibilidad casi en tiempo real, la granularidad, la asequibilidad relativa y el tamaño representan algunas de las cualidades clave que hacen que los datos del mercado laboral en línea sean especialmente adecuados para muchas formas de análisis, tanto tradicionales como novedosos.

Para un mejor aprovechamiento de los datos, se sugiere adoptar diferentes técnicas para tratar el problema de la no representatividad que incluyen enfoques estadísticos como la ponderación y herramientas aplicadas en la fase de preparación de datos (eliminación de duplicados, comparación de datos), combinándolos con enfoques de métodos mixtos que apuntan

a maximizar los datos combinando registros administrativos de oferta educativa, como análisis de encuestas de empleo, y en algunos casos específicos combinarlos con entrevistas cualitativas (como con gerentes de recursos humanos), para interpretar correctamente los sesgos.

Abordar los sesgos de los datos de vacantes de las ofertas en línea es una tarea ardua, ya que la población total de vacantes es substancialmente desconocida. No obstante, es probable que los problemas de representatividad varíen en diferentes contextos; deben evaluarse a nivel de país y con respecto a objetivos de investigación específicos, preguntas de investigación y alternativas existentes, en términos de precisión, costos y oportunidad. Cualquier análisis que utilice datos de las ofertas en línea debe integrarse en un contexto particular, y los pasos analíticos y las decisiones deben definirse y especificarse claramente. Idealmente poder contar con algoritmos de descarga, como de agrupación de ocupaciones y habilidades que deben registrarse y actualizarse para permitir la replicación del análisis.

Los datos en línea, incluidas las encuestas recopiladas en la web, representan una alternativa que abre nuevos horizontes y oportunidades en la investigación. Los datos recopilados en la web también brindan información que es difícil de recopilar por otras vías, incluso a partir de datos representativos, como los salarios (reales y esperados) o aspectos de los sesgos de género en los mercados laborales.

Es importante destacar que los principales repositorios de datos están trabajando cada vez más en colaboración con los investigadores, incluso en artículos de coautoría, en un grado mucho mayor que en otros dominios de las ciencias sociales. Como resultado, la investigación de ofertas laborales en línea es una de las áreas clave donde hay un rápido avance en la aplicación del aprendizaje automático y la inteligencia artificial en las ciencias sociales.

▶ Anexos

Anexo 1. Resumen de estudios aplicativos de *Big Data* en el mercado laboral

Sector	Escenario	País	Año	Proyecto	Fuente de datos	TÉCNICAS
Sector Público	Análisis de vacantes	Europa (28 países)	Actualización varias veces en 2021	Online Vacancy Analysis Tool for Europe (OVATE) por CEDEFOP	Portales de empleo privados, portales de servicios públicos de empleo, agencias de reclutamiento, avisos de periódicos en línea y páginas de empleadores.	<i>Web scrapping, crawling</i>
		Australia	2021	(Index Internet Vacancy) Ministerio de trabajo y pequeñas empresas	Portales de empleo privados, portales de servicios públicos de empleo, agencias de reclutamiento, avisos de periódicos en línea y páginas de empleadores.	<i>Web scrapping, crawling</i>
	Análisis de competencias	Nueva Zelanda	2021	Job vacancies (Ministerio de Empresa, Innovación y Empleo)	La tendencia en el número de anuncios de trabajo en línea por ocupación.	<i>Web scrapping, crawling</i>
	Empleo (Sector ciberseguridad)	Estados Unidos	2017-2021	CyberSeek.org incluye privados y financiación estatal	Base de vacantes, datos del gobierno, datos de certificación en orden a las ocupaciones.	<i>Web scrapping, crawling</i>
	Relación demanda laboral y oferta educativa	Uruguay	2022	Herramienta de aprendizaje automático	Microdatos oficiales, definiciones de ocupaciones, CIUO-08.	<i>Learning machine</i> (Procesamiento de lenguaje natural)
Sector privado	Demanda laboral	Global	2019	Índice de empleo en línea Universidad de OXFORD	Plataformas de trabajo en línea por país y ocupación.	
		Estados Unidos	2017	Burning Glass Technologies (servicio de paga)	Información de miles de plataformas de empleo en el país y las complementa con otras fuentes de vacantes de empleo.	
		España	Actualización anual	Observatorio de talento Bizkaia Basque (Asociativo sin lucro)	Portales de empleo, universidades, datos del gobierno.	Predictivo
	Brechas	Australia	2012-2020	University of Technology Sydney	Datos oficiales.	Predictivo <i>learning machine</i>

Sector	Escenario	País	Año	Proyecto	Fuente de datos	TÉCNICAS
Organizaciones internacionales	Formación, habilidades	UNESCO-Malawi	2018	my.job.com	Plataformas de trabajo en línea por país y ocupación.	Inteligencia artificial
	Intermediación laboral, mercado laboral	OIT Primer informe: desajustes entre demanda y oferta de trabajo, una mirada desde las ocupaciones	2021	Monitor del mercado laboral	Buscojobs y Portales web de búsqueda de empleo.	<i>Web Scraping</i>
	Brechas de habilidades, tipos de vacantes	OIT/Cinterfor (República Dominicana)	2021	OIT y OIT/Cinterfor para los Ministerios del Trabajo	Buscojobs y Portales web de búsqueda de empleo.	<i>Web Scraping</i>
	Demanda laboral y Oferta laboral	CEPAL (33 Países)	2019	Proyecto CEPAL de seguimiento de vacantes de empleo en línea en América Latina y el Caribe	Los portales seleccionados: Bumeran, CaribbeanJobs, CaribbeanJobsOnline, CompuTrabajo, Freelancer, JobisJob, Profdir y Upwork.	<i>Scrapping-Crawling</i>
		OIT/Cinterfor (Uruguay)	2019	OIT y OIT/Cinterfor para los Ministerios del Trabajo	Buscojobs y Portales web de búsqueda de empleo.	<i>Web Scraping</i>
		OIT/Cinterfor (Paraguay)	2021	OIT y OIT/Cinterfor para los Ministerios del Trabajo	Buscojobs, Paraempleo, Redtrabaje, Opción Empleo, BuscoJobs, CompuTrabajo, LinkedIn, Pivot y Redi en el periodo de junio a noviembre de 2020.	<i>Web Scraping</i>

Fuente: Elaboración propia.

► Referencias

- Aguilar, E., & Álvarez, H. (2020). *La Era de Big Data: El conocimiento es el nuevo oro*. Obtenido de <https://revistas.utp.ac.pa/index.php/el-tecnologico/article/view/439/html#:~:text=Una%20definici%C3%B3n%20v%C3%A1lida%20de%20%E2%80%9CBig,de%20decisiones%20en%20las%20organizaciones%E2%80%9D>.
- Buscojobs, CED. (2022). *Monitor Laboral: Desajustes entre oferta y demanda en el mercado laboral uruguayo: Una mirada de las ocupaciones*.
- CAF. (2017). *Lineamientos para la identificación y el cierre de brechas de capital humano para las apuestas productivas departamentales del país*. Obtenido de Banco de Desarrollo para América Latina -CAF-: <http://repositorio.esumer.edu.co/jspui/handle/esumer/1483>
- CPC-PNUD-CAF. (2019). *Lineamientos para la identificación y cierre de Brechas de Capital Humano*. Obtenido de Consejo Privado de Competitividad: <https://compite.com.co/proyecto/lineamientos-para-la-identificacion-y-cierre-de-brechas-de-capital-humano/>
- DATAFLOQ. (2019). *Big Data Analytics Paving The Path For Businesses With More Informed Decisions*. Obtenido de <https://datafloq.com/read/big-data-analytics-paving-path-businesses-decision/>
- DNP. (2016). *Big Data: Colombia entra en la revolución de los datos*. Obtenido de Departamento Nacional de Planeación: <https://www.dnp.gov.co/Paginas/Big-Data-Colombia-entra-en-la-revoluci%C3%B3n-de-los-datos-.aspx#:~:text=El%20Big%20Data%20fue%20definido,y%20el%20'software'%20tradicionales>.
- European Training Foundation. (2019). *Big Data for labour market intelligence: An introductory guide*. Obtenido de European Training Foundation: <https://www.etf.europa.eu/en/publications-and-resources/publications/big-data-labour-market-intelligence-introductory-guide>
- Gontero, S., & Novella, R. (2021). *El futuro del trabajo y los desajustes de habilidades en América Latina*. Santiago: CEPAL.
- ICO. (2018). *Big Data, artificial intelligence, machine learning and data protection*. Information Commissioner's Office.
- ILO. (2019). *Skills and jobs mismatches in low- and middle-income countries*. International Labour Office.

- ILO. (2020). *The feasibility of using big data in anticipating and matching skills needs*. Obtenido de International Labour Organization: https://www.ilo.org/skills/areas/skills-training-for-poverty-reduction/WCMS_759330/lang--en/index.htm
- Kantardzic, M. (2011). *DATA MINING Concepts, Models, Methods, and Algorithms*. http://www.ru.ac.bd/stat/wp-content/uploads/sites/25/2019/03/505_02_00_Kantardzic-Data-Mining_-Concepts-Models-Methods-and-Algorithms-Second-Edition.pdf: University of Louisville.
- McKinsey & Company. (2011). *Big data: The next frontier for innovation, competition, and productivity*. McKinsey & Company 2: https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/big%20data%20the%20next%20frontier%20for%20innovation/mgi_big_data_full_report.pdf.
- Mena Roa, M. (2021). *El Big Bang del Big Data*. Obtenido de Statista | Datos Digitales: <https://es.statista.com/grafico/26031/volumen-estimado-de-datos-digitales-creados-o-replicados-en-todo-el-mundo/>
- Ministerio de Trabajo. (s.f.). *Identificación y medición de brechas de capital humano*. Obtenido de Análisis, Monitoreo y Prospectiva Laboral: <https://www.mintrabajo.gov.co/empleo-y-pensiones/empleo/analisis-monitoreo-y-prospectiva-laboral/identificacion-y-medicion-de-brechas-de-capital-humano>
- Nexus. (2022). *Data Science vs. Big Data vs. Data Analytics*. Obtenido de <https://nexusintegra.io/>: <https://nexusintegra.io/es/data-science-vs-big-data-vs-data-analytics/>
- OIT (Vol.1). (2016). *Guías metodológicas sobre anticipación y ajuste entre oferta y demanda de habilidades: El uso de la información sobre el mercado de trabajo. Volumen 1*. Ginebra, Suiza: Copyright © Organización Internacional del Trabajo.
- OIT (Vol.2). (2016). *Guías metodológicas sobre anticipación y ajuste entre oferta y demanda de habilidades: El desarrollo de estudios prospectivos, escenarios y anticipación de competencias. Volumen 2*. Ginebra, Suiza: Copyright © Organización Internacional del Trabajo.
- OIT (Vol.2). (2017). *El desarrollo de estudios prospectivos, escenarios y anticipación de las competencias. Volumen 2*. Montevideo, Uruguay: Guías metodológicas sobre anticipación y ajuste entre oferta y demanda de habilidades.
- OIT (Vol.3). (2016). *Guías metodológicas sobre anticipación y ajuste entre oferta y demanda de habilidades: Trabajando en el ámbito sectorial. Volumen 3*. Ginebra, Suiza: Copyright © Organización Internacional del Trabajo.
- OIT (Vol.3). (2017). *Trabajando en el ámbito sectorial. Volumen 3*. Montevideo, Uruguay: Guías metodológicas sobre anticipación y ajuste entre oferta y demanda de habilidades.

- OIT (Vol.4). (2016). *Guías metodológicas sobre anticipación y ajuste entre oferta y demanda de habilidades: El papel de los proveedores de servicios de empleo. Volumen 4*. Ginebra, Suiza: Copyright © Organización Internacional del Trabajo.
- OIT. (2016). *Compendio de guías metodológicas sobre anticipación y ajuste entre oferta y demanda de habilidades*. Obtenido de Copyright © Organización Internacional del Trabajo: <https://www.oitcinterfor.org/node/7070>
- OIT. (2017). *Metodología de identificación y medición de brechas de capital humano -IMBCH-*. Bogotá D.C., Colombia: Ministerio de Trabajo: Subdirección de Análisis, Monitoreo y Prospectiva Laboral.
- OIT. (2017). *Metodología de identificación y medición de brechas de capital humano -IMBCH-*. Bogotá D.C., Colombia: Ministerio de Trabajo: Subdirección de Análisis, Monitoreo y Prospectiva Laboral.
- OIT, Vol. 4. (2017). *El papel de los proveedores de servicios de empleo. Volumen 4*. Montevideo, Uruguay.: Guías metodológicas sobre anticipación y ajuste entre oferta y demanda de habilidades.
- OIT/Cinterfor (Vol.1). (2017). *El uso de la información sobre el mercado de trabajo. Volumen 1*. . Montevideo, Uruguay: Guías metodológicas sobre anticipación y ajuste entre oferta y demanda de habilidades.
- OIT/Cinterfor. (2017). *Metodología de identificación y medición de brechas de capital humano -IMBCH-*. Bogotá D.C., Colombia: Ministerio de Trabajo: Subdirección de Análisis, Monitoreo y Prospectiva Laboral.
- PROMIDAT. (2021). *Promidat Iberoamericano*. Obtenido de Universidad Autónoma de Centro América : <https://promidat.website>
- Sonia, G., & Emanuel, M. (2021). *Macrodatos (Big data) y mercado laboral*. Naciones Unidas.
- Sosa Escudero, W. (2021). *Big Data y ciencia de datos: conceptos, oportunidades y desafíos*. Obtenido de Organización Internacional del Trabajo -OIT-: https://www.ilo.org/santiago/publicaciones/notas-informativas-cono-sur/WCMS_769307/lang--es/index.htm#:~:text=y%20desaf%C3%ADos%20...-,Big%20Data%20y%20ciencia%20de%20datos%3A%20conceptos%2C%20oportunidades%20y%20desaf%C3%ADos,aplicaci%C3%B3n%20en%20la%20
- Urquidi, M., Ortega, G., Arza, V., & Ortega, J. (2021). *Nuevas tecnologías para el empleo: Beneficios de la implementación de servicios en el marco de una arquitectura empresarial*. BID.
- Vargas, F. (2022). *Tres ideas derivadas de tres estudios de big data en el mercado de trabajo*. https://www.oitcinterfor.org/sites/default/files/file_publicacion/NOTA%2011%20FINAL%20FLIP_0.pdf: OIT/Cinterfor Notas N° 11 Febrero 2022.



Organización
Internacional
del Trabajo

CINTERFOR