

DOCUMENTO DE SÍNTESIS

PROSPECCIÓN Y DETECCIÓN DE NECESIDADES FORMATIVAS A TRAVÉS DE LA TECNOLOGÍA BIG DATA EN CASTILLA- LA MANCHA

ENTIDAD: CCOO CLM

C.I.F.: G4508033

Nº DE EXPEDIENTE: PE/2020/10

DENOMINACIÓN ACCIÓN:

Prospección y detección de necesidades formativas a través de la tecnología Big Data en CLM

TIPO DE ACCIÓN:

Prospección y detección de necesidades formativas del sistema productivo

30 de octubre de 2021

ÍNDICE

MEMORIA TÉCNICA:	4
DESARROLLO DEL PROYECTO	4
FASE 1. INVESTIGACIÓN DOCUMENTAL	5
INTRODUCCIÓN	5
I. JUSTIFICACIÓN DE LA TÉCNICA	6
II. MARCO TEÓRICO	7
II.I. DEFINICIÓN DE BIG DATA	7
II.II. EVOLUCIÓN DEL BIG DATA: DESDE EL ORIGEN DEL TÉRMINO A LA ACTUALIDAD	10
II.III. LAS 5 VS	12
II.IV. TIPOS DE DATOS	13
II.V. ALMACENAMIENTO	14
II.VI. KPI's (Indicadores Claros)	14
II.VI. CONCLUSIONES EXTRAÍDAS DEL PROYECTO "PE/2019/02: ASESORAMIENTO A PYMES SOBRE LA APLICACIÓN DE LA TECNOLOGÍA BIG DATA"	15
II.VIII. ESTADÍSTICAS DE BIG DATA	16
II.IX. COMPARATIVA A NIVEL REGIONAL Y NACIONAL	17
II.X. INICIATIVAS DE ÉXITO	20
II.XI. FASES DE UN PROYECTO BIG DATA E INDICADORES PARA LA DETECCIÓN DE NECESIDADES FORMATIVAS EN CASTILLA-LA MANCHA	24
II.XII. LIMITACIONES	27
II.XIII. EQUIPO DE PROYECTO	29
II.XIV. INDICADORES	29
II.XV. FUENTES	29
II.XVI. DATOS PRIMARIOS	30
II.XVII. ALMACENAMIENTO	30
II.XVIII. TRANSFORMACIÓN	31
II.XIX. VISUALIZAR	31
II.XX. HERRAMIENTAS QUE PUEDEN SERVIR PARA EL PROCESAMIENTO DE DATOS	32
IV. ACTIVIDADES REALIZADAS	34
FASE 2. TRABAJO DE CAMPO	36

INTRODUCCIÓN	36
I. ENTREVISTAS EN PROFUNDIDAD	36
II. GRUPO DE TRABAJO EXPERTO	40
II.I. Justificación de la técnica	40
II.II. Briefing	40
III. GESTIÓN Y PLANIFICACIÓN DEL TRABAJO DE CAMPO	41
IV. ACTIVIDADES REALIZADAS	42
FASE 3. ANÁLISIS	44
INTRODUCCIÓN	44
I. ANÁLISIS DE LOS RESULTADOS	45
<i>¿Qué aporta la tecnología big data a un proyecto como el que se postula sobre detección de necesidades formativas a través de la tecnología big data?</i>	45
<i>Fases de un proyecto de analítica de datos: Big Data</i>	46
<i>Las posibilidades del Big Data en la mejora del empleo y la formación de los trabajadores/as</i>	49
II. CONCLUSIONES	49
III. ACTIVIDADES REALIZADAS	50
FASE 4. EDICIÓN, PUBLICACIÓN Y DIFUSIÓN	52
INTRODUCCIÓN	52
I. INFOGRAFÍAS CON LAS CONCLUSIONES DEL ESTUDIO	52
II. INFORME EJECUTIVO	57
III. NEWSLETTER	60
IV. CRONOGRAMA DE DIFUSIÓN	61
V. ACTIVIDADES REALIZADAS	62
BIBLIOGRAFÍA DE TODAS LAS FASES	63
TRABAJO DE CAMPO: HERRAMIENTAS DE ANÁLISIS E INDICADORES DE VALORACIÓN METODOLÓGICA	66
RELACIÓN DE PARTICIPANTES	69
CONCLUSIONES Y PROPUESTAS DE MEJORA	77

MEMORIA TÉCNICA: DESARROLLO DEL PROYECTO

FASE 1. INVESTIGACIÓN DOCUMENTAL

INTRODUCCIÓN

El presente proyecto surge de la necesidad de desarrollo de un sistema de prospección de necesidades formativas a través del uso de la tecnología Big Data. Tras las conclusiones emergidas del estudio subvencionado por el Ministerio de Empleo y Seguridad Social y la Junta de Comunidades de Castilla-La Mancha con expediente: PE/2019/02 “*Asesoramiento a PYMES sobre la aplicación de la tecnología Big Data en la formación de sus trabajadores y la creación de empleo*”, se plantea definir las fases a seguir para el desarrollo de un sistema de tratamiento de datos a gran escala que permita detectar las necesidades formativas de los trabajadores/as de Castilla-La Mancha, al igual que las demandas de ocupaciones del mercado de la región.

La Ley 30/2015 de 9 de septiembre, por la que se regula el Sistema de Formación Profesional para el empleo en el ámbito laboral, señala en su Cap.2, Art. 4, la importancia de desarrollar un sistema eficaz de detección y prospección de necesidades formativas, que logre que estas se ajusten a los requerimientos del mercado y se anticipen a sus cambios, contribuyendo así al desarrollo profesional y personal de los trabajadores/as, y a la mejora competitiva de las empresas. Una de las principales ineficiencias identificadas en el sistema de formación profesional para el empleo, es esa escasa anticipación a las necesidades, que imposibilita una planificación de la actividad formativa ajustada a la realidad. El uso de un sistema de análisis de datos a gran escala puede ser clave para alcanzar este objetivo, y mejorar las estrategias y políticas de formación para el empleo.

La tecnología Big Data es un activo clave para la economía global. Se posiciona como una de las TICC con más proyección dada su transversalidad a todos los ámbitos. El uso de este tipo de tecnologías se extiende más allá del sector privado. La Administración Pública integra desde hace años este tipo de sistemas como herramientas útiles para la toma de decisiones y articulación de estrategias que se ajusten a la realidad. Como ejemplo de ello, el tratamiento de la crisis sanitaria del COVID-19. En el último año, se han procesado, tratado y cruzado miles de datos para estudiar la pandemia desde multitud de espacios, tanto públicos como privados, para conseguir así crear modelos predictivos fiables y nuevas herramientas frente al virus. El uso de las aplicaciones relacionadas con la analítica de datos es inmenso, ofreciendo conocimiento de valor al conectar, procesar e interpretar datos de manera continua e inmediata.

Los resultados del estudio PE/2019/02, anteriormente citado, pusieron de manifiesto la complejidad en el desarrollo técnico de un sistema que aplicase la tecnología Big Data. Por ello, a modo de conclusiones, se detallaron una serie de pasos que facilitarían su abordaje. Uno de los apartados de esta primera fase incluirá y describirá el desarrollo de estos a través de la información documental abordada.

Todo ello lleva a plantear un proyecto que permita definir de manera precisa todos los aspectos de creación de un sistema de analítica de datos a gran escala que

consiga cumplir con los objetivos previstos de prospección y detección de necesidades formativas, y que proporcione un mejor conocimiento de las cualificaciones y perfiles profesiones demandados, para de esta manera conseguir una mayor sincronía entre oferta y demanda formativa, facilitando la creación de estrategias de formación acordes a las demandas de los/as trabajadores/as y el mercado, y adaptadas a la realidad regional de Castilla-La Mancha.

La recesión económica derivada del impacto del Covid-19 ha afectado al mercado de trabajo de manera desigual, afectando en mayor medida a las actividades de hostelería, transportes, y turismo, y entre los trabajadores/as, a los asalariados con contratos temporales. Los rasgos estructurales que caracterizan el empleo en España se caracterizan por una tasa de paro elevada (sobre todo entre los jóvenes), alta temporalidad y precariedad laboral. Al mismo tiempo, los datos arrojan que las personas menos cualificadas son las más damnificadas por la pandemia. Estos hechos, unidos al nuevo paradigma tecnológico que cambiará las capacitaciones demandadas en los puestos de trabajo por las empresas, instan a poseer un conocimiento exhaustivo que permita anticiparse a las tendencias y necesidades futuras, y convierta a la formación para el empleo en una palanca para la recuperación y reconversión económica tanto de Castilla-La Mancha, como del país.

I. JUSTIFICACIÓN DE LA TÉCNICA

Inicialmente, para contextualizar la temática del proyecto, se llevará a cabo una revisión bibliográfica. Esta técnica de investigación, es una modalidad de trabajo que tiene como fin recopilar información de fuentes secundarias (estudios, estadísticas y demás fuentes de datos), que permitan conocer el estado de la cuestión y los posibles enfoques para la investigación, así como las reflexiones precedentes sobre el objeto de estudio.

Los objetivos a cumplir en esta primera fase de investigación documental son:

- Analizar datos de fuentes secundarias, informes, estadísticas, estudios actuales, que den contexto para enmarcar la tecnología Big Data a nivel nacional y regional
- Identificar iniciativas de uso de la tecnología Big Data
- Recopilar datos sobre las características técnicas de cada una de las fases necesarias para la implantación de un proyecto que requiera de la tecnología Big Data
- Revisar bibliografía sobre los posibles indicadores y/o variables que faciliten la detección de necesidades formativas

En el momento de seleccionar las fuentes, se tiene que tener en cuenta aspectos como: calidad de las mismas, idoneidad según el tema propuesto, relevancia y utilidad. Además, el volumen de información se debe controlar para no paralizar el

desarrollo de ideas, y es útil contar con un orden correcto de los recursos utilizados para su referencia y/o consulta.

Los principales puntos a seguir, representados sintéticamente son (Barry, 1999, ACRL/ALA 2000):

- Formulación, análisis y clarificación de la necesidad de información
- Identificación, observación, selección y rechazo de fuentes
- Localización y acceso a la información
- Extracción, registro y gestión de la información
- Evaluación de la información
- Integración del conocimiento adquirido
- Uso de la información para el objetivo de la investigación
- Comunicación de la información
- Actuación ética en el acceso y uso de la información

En conclusión, la elaboración de esta primera fase establece un contexto de partida, que aclara y expone los procesos propios de la tecnología Big Data, que, dada su tecnicidad, están dotados de cierta complejidad. Este sistema Big Data se concibe para lograr una mayor detección de las necesidades formativas del tejido productivo castellanomanchego. Por otro lado, el análisis documental sirve de base para el diseño metodológico posterior: elaboración de guías de entrevistas y definición de perfiles de los informantes clave que puedan dar respuesta a los objetivos previstos en el proyecto.

II. MARCO TEÓRICO

II.1. DEFINICIÓN DE BIG DATA

El término *Big Data*, de origen inglés, se traduce al español como “macro datos”, es decir, datos a gran escala. Esta sería la definición básica, su traducción literal. Sin embargo, la expresión *Big Data*, hace referencia a multitud de acepciones, dependiendo de quién la formule. Se trata de un concepto empleado tanto para hablar de datos por sí mismos, como para hacerlo sobre los métodos de análisis destinados a su explotación (*predictive analytics, data mining, big data analytics, machine learning, neural networks...*). Algunos de los propósitos principales de estos procesos son: identificación de patrones, modelos predictivos, apoyo al diagnóstico y/o toma de decisiones.

El catedrático Gastón Becerra, acerca e interpreta la siguiente definición sobre el término de Gandomi & Haider en su artículo “Big data como objeto de estudio y método para la investigación empírica en sociología y psicología social” : *“Comúnmente, se lo suele caracterizar (al Big Data) como un fenómeno tecnológico en el que confluyen la disponibilidad de grandes volúmenes de datos en muy diversos formatos, la proliferación de nuevas técnicas de análisis que permiten un*

procesamiento veloz de los mismos, y el desarrollo de una infraestructura de sistemas capaz de soportarlos (Becerra, G., 2018). Esta definición, da una idea de la complejidad que posee el abordaje del término.

A continuación, se muestran una serie de definiciones que se pueden encontrar tanto a nivel institucional, como de publicaciones oficiales o entre el personal de investigación:

“Conjunto de técnicas que permiten analizar, procesar y gestionar conjuntos de datos extremadamente grandes que pueden ser analizados informáticamente para revelar patrones, tendencias y asociaciones, especialmente en relación con la conducta humana y las interacciones de los usuarios” (Real Academia Española, 2019).

“La gestión y análisis de enormes volúmenes de datos que no pueden ser tratados de manera convencional, ya que superan los límites y capacidades de las herramientas de software habitualmente utilizadas para la captura, gestión y procesamiento de datos” (El economista, 2014).

“Hoy en día el término “big data” es una constante y, aun así, la mera noción sigue generando confusión. Esta palabra se ha utilizado para trasladar al público todo tipo de conceptos entre los que se incluyen grandes cantidades de datos, analítica de redes sociales, herramientas de última generación para gestionar los datos, datos en tiempo real y mucho más” (IBM, 2012).

“Los ‘big data’, los datos masivos, se refieren a cosas que se pueden hacer a gran escala, pero no a una escala inferior, para extraer nuevas percepciones o crear nuevas formas de valor, de tal forma que transforman los mercados, las organizaciones, las relaciones entre los ciudadanos y los gobiernos, etcétera” (Mayer-Schönberger, 2013).

Todas estas definiciones dotan de gran amplitud al término y en consecuencia de un mayor alcance, pero al mismo tiempo restan concreción, pues hacen referencia también a su posterior explotación y análisis, lo que lo transforma en cierta manera en inabarcable.

Para otras muchas voces autorizadas, el concepto *Big Data* se asocia exclusivamente a grandes cantidades de datos que tienen como característica intrínseca su gran volumen, variedad de forma y velocidad, alejándose de las concepciones anteriores que englobaban su posterior tratamiento.

“Fundéu: “Macrodatos e inteligencia de datos son alternativas en español a la voz inglesa big data, que se emplea en el sector de las tecnologías de la información y de la comunicación para aludir a un conjunto de datos que, por su volumen y variedad y por la velocidad a la que necesitan ser procesados, supera las capacidades de los sistemas informáticos habituales. El término macrodatos es una alternativa válida pues aporta, como big, el significado de ‘grande’; es una solución breve y no tiene, como ocurriría con megadatos, un posible riesgo de confusión con el término mega, muy frecuente también en los mismos ámbitos” (Fundae, 2019).

“Los datos masivos, también llamados macrodatos o Big Data, son grandes cantidades de información que por su volumen, variedad y velocidad de obtención requieren tecnologías y métodos especializados para su aprovechamiento” (Entinema, 2017).

“Big Data son datos que contienen una mayor variedad y que se presentan en volúmenes crecientes y a una velocidad superior” (Gartner, 2001).

Tomando en consideración el objetivo general del presente estudio, y dado las características del mismo, se utilizará la idea de estas últimas definiciones para la operacionalización del término: gran volumen de datos. Entendiendo entonces, la tecnología *Big Data* como la expresión que abarca todos los procesos y análisis posteriores que se adaptarán a las necesidades del objeto de estudio.

Es decir, Big Data, como gran volumen de datos y tecnología Big Data como todas las acciones posteriores de tratamiento de datos.

Dentro de estos tratamientos posteriores, existen diferentes fases de desarrollo y herramientas de tratamiento de datos *“advanced analytics: predictive analytics, data mining, big data analytics, machine learning, neural networks,...* que se emplean fundamentalmente para la identificación de patrones o el desarrollo de modelos predictivos y, también, a los desarrollos de inteligencia artificial: *machine learning, deep learning, natural language processing, cognitive computing, ...*” (Barea, J. et al., 2018) Estos variarán según el objetivo final que la entidad o agente tenga, y las aplicaciones que quieran dar a tal información. En este caso, sería conocer las necesidades demandadas y requeridas tanto por los trabajadores/as, como por los agentes formadores y el mercado laboral. Sería necesario conocer las fuentes disponibles, las herramientas que requieren los procesos técnicos y demás requerimientos de esta tecnología.

La necesidad de definir claramente los conceptos nace de separar la “materia prima”, los datos, de su explotación y modelación posterior. El *Big Data* no explica nada por sí mismo, son cantidades ingentes de datos agrupados. Es necesario una explotación diseñada para dar respuesta a los objetivos específicos que se planteen (KPI's) para dar a luz un conocimiento de valor. Ahí aparece la expresión, tecnología *Big Data*.

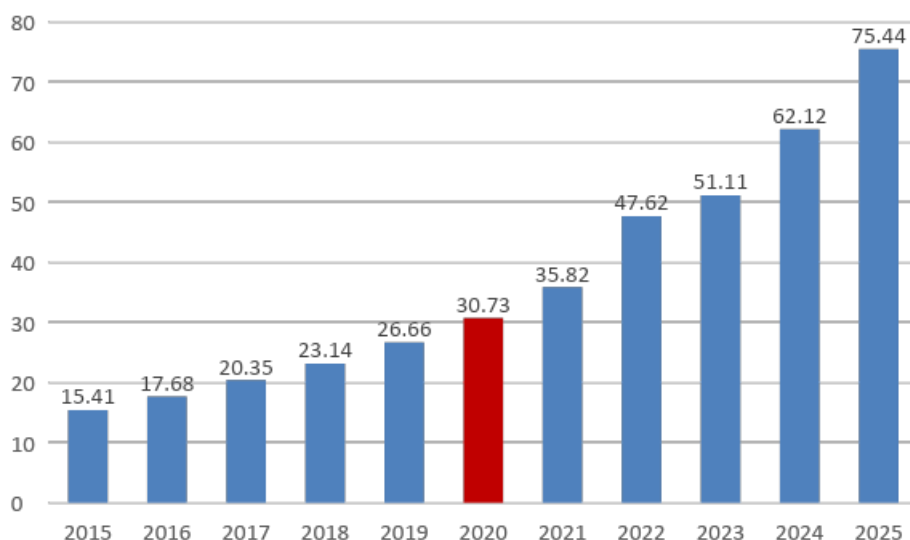
El fin de este estudio, es desengranar y detallar el procedimiento para poder crear una plataforma operativa basada en la tecnología *Big Data*, que recupere datos de formación y mercado de trabajo de Castilla-La Mancha, de manera que sea posible anticiparse a las necesidades de cualificación de las personas trabajadoras y mejorar la planificación de la FPE, y con ello, la productividad y competitividad del tejido empresarial de la región.

II.II. EVOLUCIÓN DEL BIG DATA: DESDE EL ORIGEN DEL TÉRMINO A LA ACTUALIDAD

El auge del *Big Data* es debido al aumento de la recolección de datos, gracias tanto a las posibilidades de nuevos productos electrónicos del mercado, como a las mejores capacidades técnicas de procesamiento y almacenamiento de estos datos. El mayor acceso se da por medio de redes sociales, emails, encuestas, blogs, señales GPS, apps... que están presentes en multitud de objetos de la vida cotidiana (móviles, relojes, aparatos de cocina, aspiradoras, domótica general...). E IoT, Internet de las cosas, alude a estos objetos conectados e interconectados, generando datos de manera continua y enviándolos a una nube de almacenamiento en la red.

El siguiente gráfico muestra la evolución de dispositivos conectados al Internet de las Cosas (IoT por el término inglés, "Internet of Things"). En diez años, se puede apreciar que la previsión es que se quintuple el número de dispositivos. Más del doble que en la actualidad (2020).

GRÁFICO 1. Evolución dispositivos conectados al IoT (2005-2025)



Fuente: Elaboración propia a partir de datos extraídos de Statista.

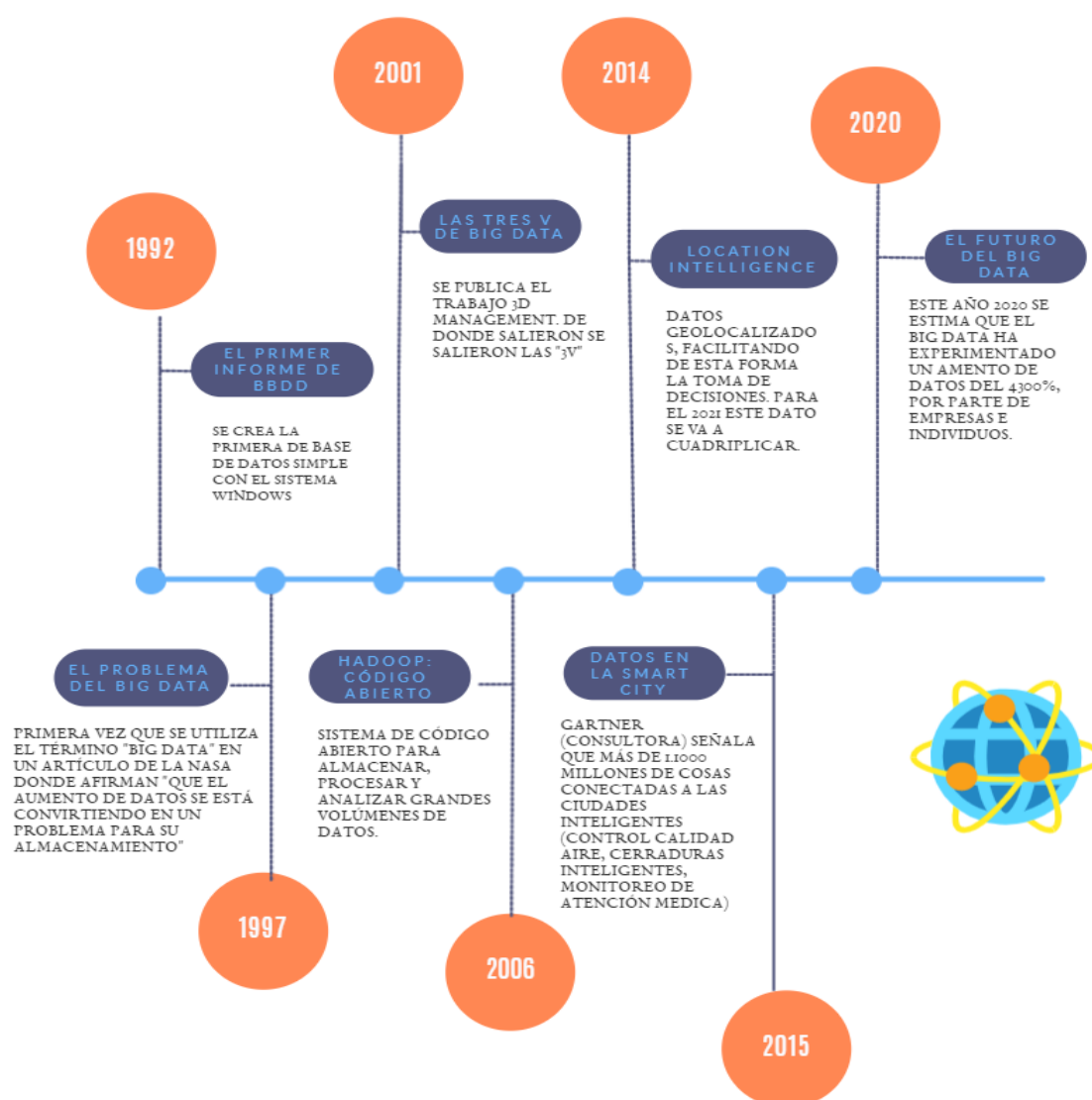
Esta representación del evolutivo, sirve para, de forma visual, mostrar las posibilidades y el gran cambio que se está produciendo dentro del Internet de las Cosas, lo que va unido a la creación, gestión y almacenamiento de datos. El crecimiento es exponencial.

La historia del Big Data no tiene un origen concreto. Los datos se han recopilado desde siempre, ya fuera mediante escritura, dibujo y/o posteriormente audio y video. A partir de la creación de los ordenadores, las grandes empresas poseían BBDD de clientes e información interna, de manera estructurada, con los que podían operar a través de estadísticas. Pero el desarrollo y avance de la tecnología, ha ido modificando los procesos de recogida y modificación de los datos hasta

llegar a la tecnología Big Data que se conoce hoy en día. Internet fue un punto de inflexión, porque una cantidad muy elevada de información se hizo accesible a través de la red, tanto para su consulta como para su estudio. Otro punto vital, fue la capacidad de almacenamiento. A medida que aumentaba el peso y volumen de los datos, lo hacía la capacidad para guardarlos.

A continuación, se muestra un “timeline” que detalla de manera sintética los momentos clave de esta tecnología. Desde la primera vez que se utilizó el término, hasta las estimaciones de uso del momento actual.

IMAGEN 1. Timeline: Evolución del Big Data



Fuente: Elaboración propia a partir de datos de “Vodafone: Ideas para tu empresa”

Una de las razones del gran avance de la *tecnología Big Data* ha sido el aumento en el uso masivo de dispositivos portátiles con conexión a la red. Según la consultora Gartner, en 2020, 22.000 millones de “cosas” estaban conectadas a internet. Por otro lado, **la mejora en las capacidades de almacenamiento** fue básica para que este crecimiento se produjese. De estos aspectos adheridos a la tecnología Big Data, se detallarán sus características en apartados posteriores. Como dato de interés, en España se encuentran 36 centros de recogida de información de los más de 2.200 centros existentes a nivel mundial.

II.III.LAS 5 VS

Al realizar la investigación documental, uno de los principales componentes teóricos acerca del Big Data son las dimensiones que componen esta tecnología, lo conforman: volumen, velocidad, variedad, valor y veracidad, en el formato clásico; al que algunos expertos añaden visualización y viabilidad.

IMAGEN 2. Las 5Vs



Fuente: “Qué es el Big Data” por Aura Quantic

Volumen: Intrínseca al concepto. Grandes cantidades de datos generados gracias a los avances tecnológicos y su automatización. A medida que se producen los avances, se crean retos técnicos para su gestión, almacenamiento...

Velocidad: La velocidad de datos ha ido creciendo, y en estos momentos, es a tiempo real gracias a la ubicuidad de los dispositivos conectados a la red. Datos actualizados y disponibles inmediatamente tras su “creación”.

Variiedad: Dependiendo de la fuente de la información, el formato de los datos es variable, en tipo y estructura. Se pueden encontrar tanto BBDD en excel, como imágenes, coordenadas GPS, incluso “rostros” (con el reconocimiento facial) .

Veracidad: Tras la recogida de datos se debe realizar un depurado de información, separando la información útil del ruido.

Valor: Si se tratan los datos , depuran, y procesan correctamente. Los datos “poseerán” información de valor, para la empresa/entidad que la requiera. Es la dimensión “más relevante” , de ahí se extraerán las decisiones y estrategias óptimas para los objetivos de la investigación.

Las dos dimensiones a mayores, que para algunos expertos también son de vital de importancia:

Variabilidad: La capacidad que contengan las estructuras creadas dentro del Big Data, para comprender los contextos de los datos y su significado, de manera que se realice un análisis semántico.

Visualización: Finalmente, para la comprensión de la información , se realizan visualizaciones que permiten de manera sencilla entender el valor de los datos rescatados.

II.IV. TIPOS DE DATOS

Los datos, la materia prima fuente de información, poseen una estructura diferente dependiendo de la fuente de la que sean extraídos.

- **Datos estructurados:** Datos que se encuentran estructurados de una manera ordenada y definida, suelen mostrarse en filas y columnas. Su procesamiento posterior es más sencillo. Como ejemplo de ellos: hojas de excel, csv, bases de datos de empresas, encuestas y formularios de páginas web..

- **Datos semiestructurados:** No es posible gestionar los datos de manera estandar. No tienen una organización 100% definida, pero si ciertamente identificable. Como ejemplo de ellos encontramos: html, json y xml, lenguajes de programación de páginas web.

- **Datos no estructurados:** Se tienen que identificar para almacenar y ordenar. No tienen una estructura identificable. Como ejemplo de ello encontramos: imágenes, videos, audios...

Este punto es crucial para llevar a cabo un proyecto Big Data: conocer la estructura de los datos para que su integración y tratamiento sea correcto, y se pueda extraer información de valor.

II.V. ALMACENAMIENTO

No solo es necesario almacenar los datos extraídos de manera correcta, también tiene que tener una seguridad eficaz. Así, las soluciones de almacenamiento deben ser eficientes, eficaces y proporcionar las siguientes características (Byte,2019):

- Fácil acceso: Gasto mínimo de tiempo de configuración y gestión
- Alta escalabilidad: Posibilidad de invertir en su crecimiento de forma rápida y sencilla para no interrumpir el servicio
- Disponibilidad: Funcionalidades que aporten “*fiabilidad, como la redundancia de hardware, flexibilidad en los planes de protección de datos y las soluciones de servidores de conmutación por error*”.

II.VI.KPI's (Indicadores Claros)

Por otra parte, *la tecnología Big Data* necesita conocer muy bien las respuestas a las siguientes preguntas, este sería el primer paso a realizar, y la clave para explotar este tipo de tecnología.

1. *Qué datos se necesitan para alcanzar los objetivos: detectar las necesidades formativas de los/as trabajadores/as de Castilla-La Mancha*

En este caso se plantea capturar datos externos e internos sobre empleo que contengan las cualificaciones y competencias que se requieren para el puesto.

2. *Qué datos existen y dónde están o se pueden conseguir (fuentes)*

Conocer y seleccionar las fuentes de donde extraer los datos. Como ejemplo, se toma la iniciativa del Inaem de prospección del mercado de trabajo y su acuerdo con Infojobs para extraer datos de su plataforma.

Se necesita una orientación técnica para responder a los siguientes puntos técnicos.

3. *Cuáles son válidos y cómo se interpretan.*
4. *Cuál es y cómo se obtiene su valor.*
5. *Cómo se puede distribuir y consumir el valor generado.*
6. *Cómo se custodia, administra y gobiernan correctamente los datos y la información.*

Como se puede apreciar, son muchas las cuestiones a tener en cuenta en la aplicación de *la tecnología Big Data*. Siendo esta una herramienta de gran potencial, hay que organizar de manera ajustada el *planning* y los recursos necesarios para obtener los resultados óptimos que la entidad requiera.

II.VI.CONCLUSIONES EXTRAÍDAS DEL PROYECTO “PE/2019/02: ASESORAMIENTO A PYMES SOBRE LA APLICACIÓN DE LA TECNOLOGÍA BIG DATA”

La crisis sanitaria derivada del Covid-19 ha acelerado la integración de la digitalización en todos los espacios de la sociedad: personal, laboral, formativo, etcétera. Floreciendo la necesidad de fomentar la cultura digital en Castilla-La Mancha, sobre todo en aquellas entidades que por la falta de recursos o por la falta de información se estén quedando atrás, como es el caso de algunas pymes. Además de esta avanzada digitalización, la reestructuración de la economía es un hecho tras el varapalo económico derivado de la pandemia. Por ello, más que nunca, se hace necesario ajustar las formaciones para que las personas trabajadoras, posean las competencias necesarias de adaptación laboral. Una solución propuesta y avanzada en este proyecto, es la creación de un sistema Big Data que sea capaz de detectar las necesidades del mercado tras su monitorización, de manera que sea sencillo determinar las formaciones según territorio y sector económico, ocupación.

El Big Data es un concepto que resuena de manera significativa en el entorno digital desde hace años. La aplicación de estos sistemas de inteligencia artificial y analítica de datos son beneficiosas para cualquier empresa o entidad que haya detectado necesidades en su centro, pero su aplicación debe ir respaldada por un equipo técnico profesional de expertos y expertas que asesore y oriente sobre la aplicación del método idóneo para, a través del valor del dato, cubrir los objetivos que se han determinado a priori. Es importante que este sea un buen asesoramiento y que analicen de forma adecuada las necesidades de la entidad para establecer un trabajo analítico conciso que no genere costes adicionales.

El procedimiento adecuado a seguir en procesamiento de datos es el siguiente:

Paso 1. Definición de las necesidades y las metas de la empresa o la entidad

Paso 2. Establecer un equipo de trabajo con expertos/as y/o se contratará una empresa consultora especializada

Paso 3. Determinar los datos a integrar

Paso 4. Depurar y preparar los datos

Paso 5. Establecer la inteligencia artificial más adecuada

Para saber si el Big Data es el método más adecuado a efectos de este estudio, es imprescindible abordar el **Paso 1 y definir a priori cuál es la necesidad a la que se quiere dar respuesta y cuáles son las metas a cubrir. En este caso la meta general, siguiendo los objetivos del proyecto, es la anticipación de las tendencias del mercado y las necesidades formativas de los/las trabajadores/as para favorecer la empleabilidad en Castilla-La Mancha.**

El cambio de escenario provocado por la crisis del covid-19 hace necesario indagar en las nuevas tendencias del mercado impulsadas, en parte, a partir de la integración forzosa de la digitalización, para garantizar la adaptación de la población a los nuevos escenarios laborales. Así como indagar en las necesidades

formativas de la población para favorecer la sincronía entre las nuevas demandas formativas de las empresas y la formación ofertada a los trabajadores/as.

Plantearse las necesidades de recualificación de personas, de perfiles que no se puedan reincorporar al mercado de forma inmediata, y redirigirlas a otros ámbitos basándose en sus capacitaciones, habilidades y experiencia es prioritario en un momento como el actual para favorecer la empleabilidad.

Los datos a integrar para abordar estas necesidades deben proceder de diversas fuentes de información y se corresponderán con ofertas formativas, demandas de empleo, información sobre la formación adquirida por la población, etcétera. Existen multitud de datos sobre formación y ofertas de empleo, que, puestos en común, darían información muy valiosa para ser más eficientes en el uso de los recursos públicos y políticas, y ajustar los planes de formación a las necesidades detectadas tras la inteligencia artificial.

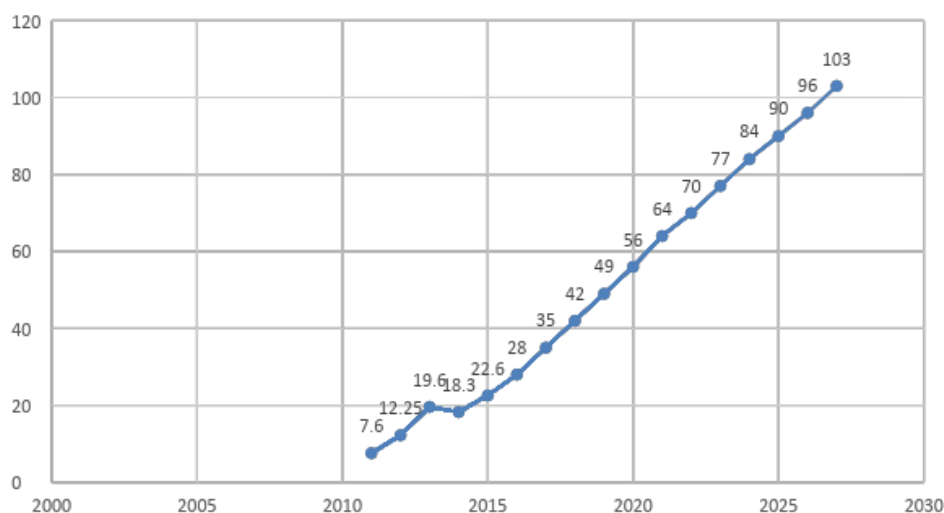
Por ello, se ha determinado la necesidad de contar con un acuerdo marco entre agentes de la administración pública, agentes sociales, y las entidades de formación ligadas a ellos, que recoja los términos y condiciones en los que se plantea el equipo de trabajo. Una vez establecido este grupo de trabajo, será más fácil acceder y juntar los datos de las diversas entidades. Desde Fundae se ofrece, por ejemplo, poner a disposición la oferta de formación a quien tiene que ir dirigida para los próximos empleos que se van a generar, o para re cualificar a esos trabajadores/as.

Este equipo de trabajo deberá contar con el asesoramiento de un equipo técnico, especializado y ayudará en la toma de decisiones sobre qué fuentes de datos se van a utilizar para resolver las necesidades y los KPIS establecidos. Además, se encargará de los próximos pasos a seguir tales como la depuración y limpieza de los datos, la selección y aplicación del mejor método analítico para su tratamiento, etcétera.

II.VIII. ESTADÍSTICAS DE BIG DATA

La primera de esta serie de estadísticas, refleja la previsión de que los ingresos del mercado de Big Data en todo el mundo para software y servicios. Se prevé que aumenten de los 56.000 millones de dólares en 2020 a los 103.000 millones en 2027. La evolución tiene un crecimiento lineal (4x) desde 2014. Este gráfico refleja el gran peso que posee esta tecnología de forma global.

GRÁFICO 2. Ingresos mundiales servicios Big Data



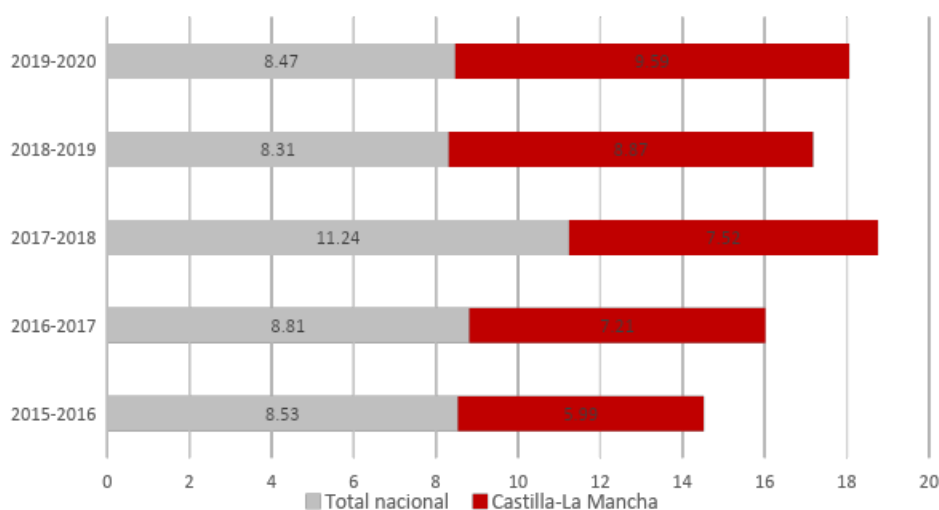
Fuente: Elaboración propia a partir de datos de Statista.

II.IX. COMPARATIVA A NIVEL REGIONAL Y NACIONAL

A través de la encuesta “Encuesta sobre el uso de las tecnologías de la información y las comunicaciones TIC y el comercio electrónico en las empresas”, se han rescatado estadísticas que muestran las tendencias de uso de estas tecnologías tanto a nivel nacional, como regional.

Los resultados seleccionados de análisis de Big Data son los correspondientes con las TIC en las empresas con 10 o más empleados (5.402 empresas en 2020 según el Instituto Nacional de Estadística)

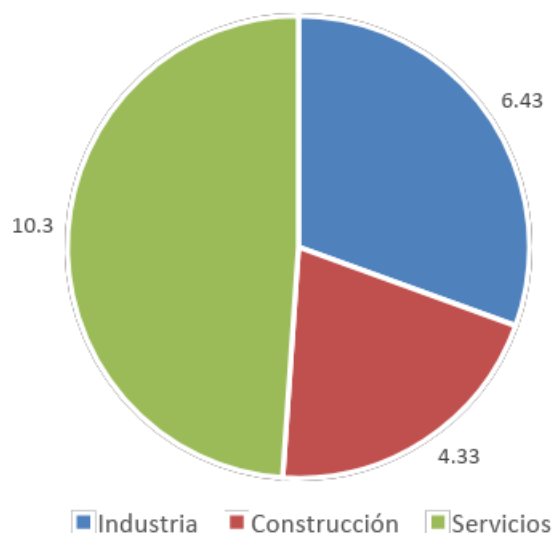
GRÁFICO 3: Empresas que analizaron grandes fuentes de datos (Big Data)



Fuente: Elaboración propia a partir de datos del INE (2020)

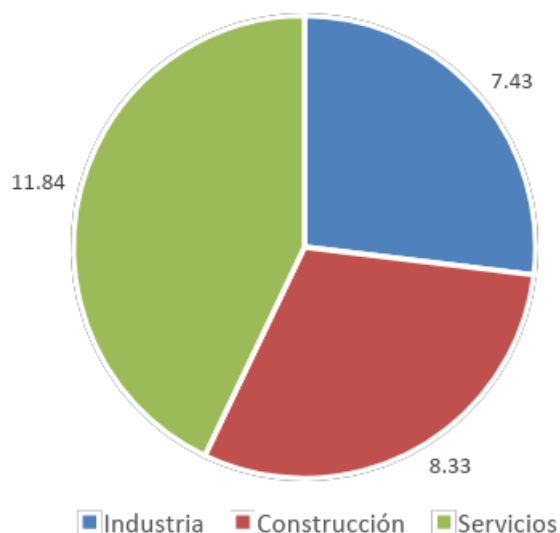
El gráfico inmediatamente superior, muestra la comparativa entre Castilla-La Mancha y España a lo largo de los años del uso del Big Data de las empresas con más de 10 empleados. Se puede observar que, aunque la diferencia ha variado en los intervalos de dos años, el último estudio muestra que Castilla-La Mancha tiene una tasa de análisis mayor que la media española.

GRÁFICO 4: Uso del Big Data por las empresas por sectores a nivel nacional – periodo 2020



Fuente: Elaboración propia a partir de datos del INE (2020)

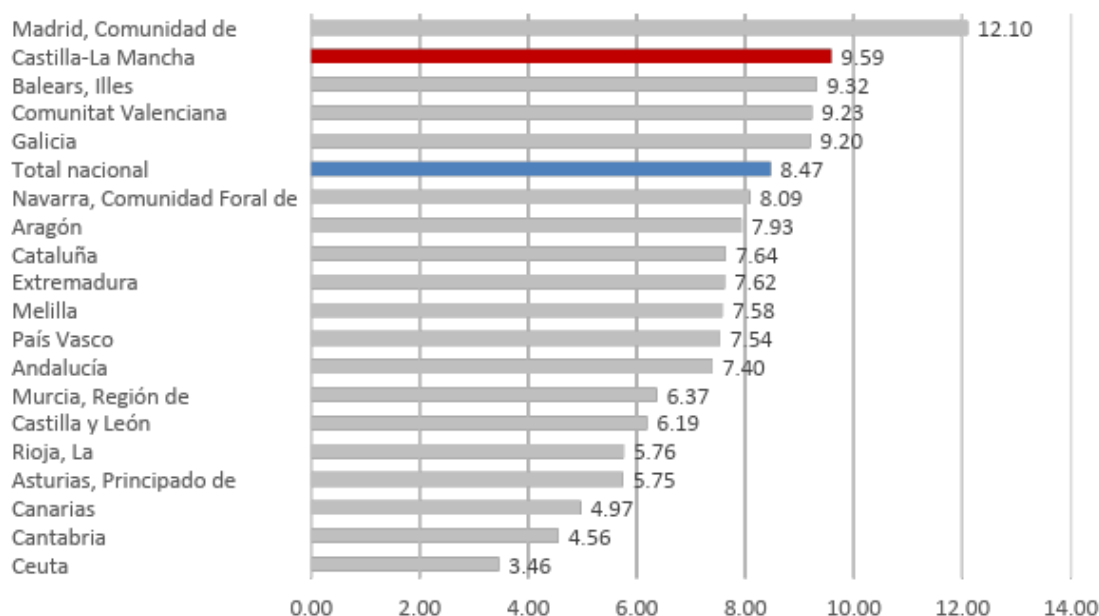
GRÁFICO 5: Uso del Big Data por las empresas por sectores a nivel Castilla-La Mancha año – periodo 2020



Fuente: Elaboración propia a partir de datos del INE (2020)

Estos gráficos circulares muestran el uso del Big Data según sectores, el superior a nivel nacional, y el inferior a nivel regional. Se observa que, en ambos casos, el sector servicios es el que integra el sistema Big Data en mayor medida.

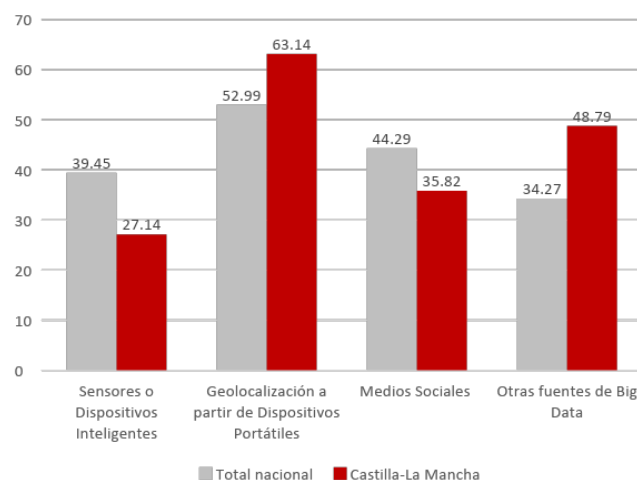
GRÁFICO 6: Uso del Big Data por año 2020 por comunidades autónomas



Fuente: Elaboración propia a partir de datos del INE (2020)

Tal como mostraba el gráfico 2, este descriptivo de barras refleja que la región castellanomanchega tiene un uso del sistema Big Data mayor que la media nacional.

GRÁFICO 7: Tipo de fuentes de datos a nivel nacional y regional, Castilla-La Mancha

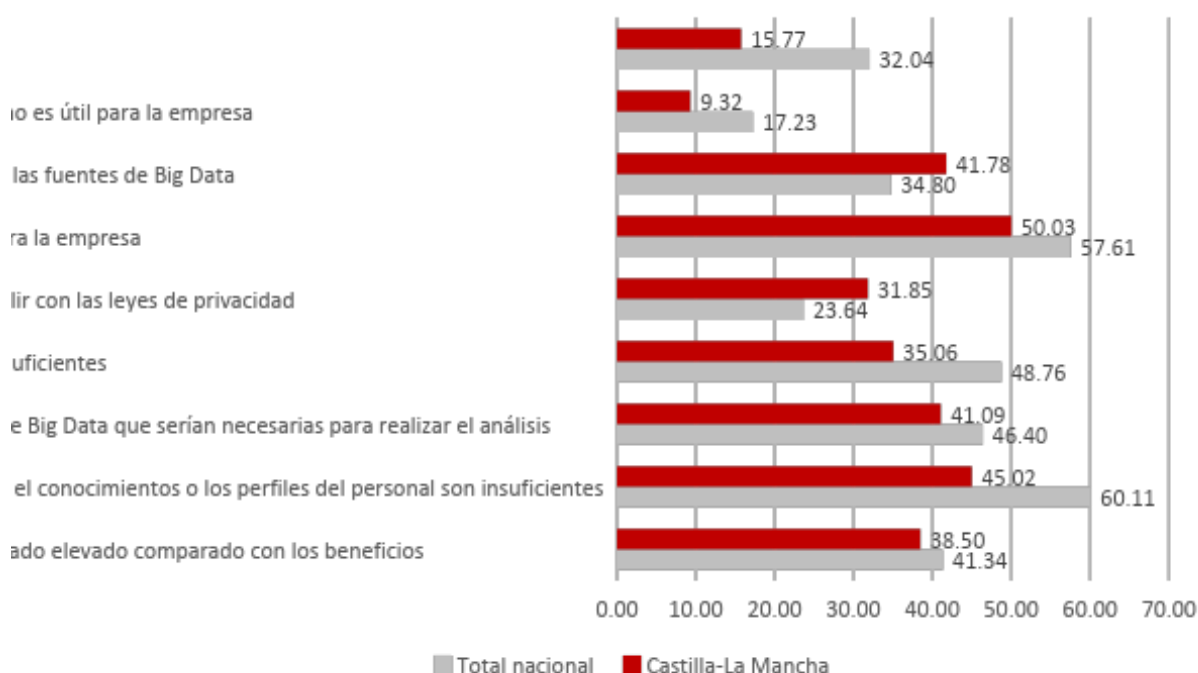


Fuente: Elaboración propia a partir de datos del INE (2020)

El gráfico inferior muestra los motivos por los que el Big Data no se integra en las empresas de más de 10 empleados de Castilla-La Mancha y España. Regionalmente las razones que destacan son:

- Una calidad insuficiente en las fuentes de Big Data
- No es una prioridad para la empresa
- Los recursos humanos, el conocimiento o los perfiles del personal son insuficientes.

GRÁFICO 8: Motivos por los que las empresas **no usan** el Big Data a nivel nacional y regional, Castilla-La Mancha



Fuente: Elaboración propia a partir de datos del INE (2020)

II.X. INICIATIVAS DE ÉXITO

ESPAÑA Y BIG DATA

Big data aplicado a la movilidad

El Ministerio de Transportes, Movilidad y Agenda Urbana posee un estudio de actualización diaria sobre la movilidad en todo el territorio español a través del Big Data. Este proyecto tiene como objetivos: proporcionar información de la movilidad a nivel nacional, de comunidad autónoma, provincial y local, así como

apoyar las labores de seguimiento de la evolución de la enfermedad, para evaluar la efectividad de las medidas de restricción de la movilidad adoptadas, y finalmente, apoyar la toma de decisiones durante esta etapa de pandemia por la COVID-19.

El acceso a sus resultados es abierto. El periodo de respuesta de la recogida y explotación de datos es de 3 días.

Nota: En 2018 ya se llevó a cabo un trabajo similar utilizando la tecnología Big Data, sobre la movilidad interprovincial a nivel nacional.

<https://www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data>

Big Data aplicado a la formación y al mercado de trabajo

El ámbito de la formación, tal como está estipulado en el proyecto, es un espacio idóneo para la aplicación de la tecnología Big Data. La Fundae analiza desde su sitio web las posibilidades que este sistema de datos ofrece. Señalan que gracias al Big Data *“se podrán identificar patrones de comportamiento en los estilos de aprendizaje permitiendo así dotar de mayor adaptabilidad y personalización al proceso formativo”* (Fundae, 2019). Además, se indica que este tipo de técnicas, ayudarían al desarrollo de una evaluación e interacción con los estudiantes más satisfactoria, ajustando las necesidades de los alumnos, con la oferta que se da. Además, se están probando técnicas que permitan analizar la formación bonificada, a cuyos datos Fundae tiene un acceso continuo, de modo que supongan un indicador del cual partir para conocer las necesidades de las empresas. Estos proyectos, sirven como premisa del objeto del presente proyecto.

Otra de las iniciativas existentes, es la aplicación del Big Data vinculado al mercado de trabajo. Uno de los proyectos creados por el Centro Europeo para el Desarrollo de la Formación Profesional tiene como fin ofrecer información sobre las competencias que se requieren en el mercado a los trabajadores/as que buscan empleo, a través del análisis de más de 30 millones de ofertas de empleo en su publicación. El estudio denominado, “The online job vacancy market in the EU”, extrajo las ofertas publicadas en la segunda mitad de 2018, procedentes de siete países europeos cuyos mercados de trabajo suponían dos tercios del empleo de la UE: República Checa, Alemania, España, Francia, Irlanda, Italia y Reino Unido. Las ofertas se obtuvieron de varias fuentes: el 79 % de portales privados de búsqueda de empleo (Indee, LinkedIn, Infojobs), el 15 % de portales públicos de empleo y el 6 % restante de otras fuentes como los periódicos digitales. Posteriormente, para su análisis, se utilizaron técnicas de Big data y machine learning (aprendizaje automático sin ser expresamente programados para ello a través de sistemas informáticos) capaces de tratar grandes volúmenes de texto en diferentes idiomas para extraer y clasificar información de estas ofertas y poderlas así analizar para obtener información de valor (en términos técnicos este proceso se denomina *scrapear*). Estas técnicas se utilizaron en combinación con la Clasificación Europea de Habilidades, Competencias, Cualificaciones y Ocupaciones (ESCO), para conocer las

competencias profesionales solicitadas en las ofertas de empleo. Los resultados se presentan en la herramienta online de análisis de las competencias en las ofertas de empleo de Europa (Skills OVATE).

Fuente: <https://blog.fundae.es/2019/06/26/big-data-aplicado-formacion-empleo/#more-2461>

<https://www.cedefop.europa.eu/en/data-visualisations/skills-online-vacancies>

<https://www.cedefop.europa.eu/en/publications-and-resources/publications/5572>

Big Data en el ámbito de la salud: lucha contra el coronavirus

1. Big Data para la creación de Apps de rastreo

El papel de las aplicaciones de rastreo del virus COVID-19 ha sido vital en muchos territorios, permitiendo:

- Notificar al usuario si ha estado en contacto con un positivo
- Identificar las zonas con mayor tasa de contagios evitando el riesgo de los usuarios

País	App	Descargas	% de descargas	Fecha de lanzamiento	Casos totales de COVID/ 1 mill po
Irlanda	COVID Tracker	1.28mill	26%	7 de julio de 2020	7,784
Alemania	Corona-Warn App	18.4 mill	22%	16 de julio de 2020	3,616
Reino Unido	NHS COVID-19	12.5 mill	19%	24 de septiembre de 2020	7,584
Portugal	StayWay Coid	1.26 mill	12%	1 de septiembre de 2020	7,768
Italia	Immuni	7 mill	12%	15 de junio de 2020	5,42
Austria	Stopp Corona	1 mill	11%	25 de marzo de 2020	5,421
España	Radar COVID	4.6 mill	10%	15 de septiembre de 2020	10,239
Bélgica	Coronalert	0.65 mill	5,50%	30 de septiembre de 2020	11,224
Francia	StopCovid	2.5 mill	3,50%	02 de junio de 2020	9,481

Fuente: Elaboración propia a partir de datos extraídos de Selecta

2. Big Data en la investigación contra la Covid-19.

Existen, a día de hoy, cientos de estudios que analizan el virus para conseguir entender su comportamiento. Según la revista Science¹, desde enero de 2020, se han publicado más de 23.000 estudios científicos sobre la Covid-19 y sus efectos en la salud de las personas.

3. Big Data en el estudio del comportamiento del virus y la evolución de la enfermedad.

Gracias a la recopilación continua de datos, y su procesamiento posterior a través de potentes dispositivos, es posible predecir y desentrañar características del Covid-19. Estos resultados fueron y son utilizados para el diseño de fármacos que combaten el virus, del mismo modo que para la creación de las vacunas. Como ejemplo de aplicación, un grupo de investigadores de la Universidad Politécnica de Valencia (UPV) ha creado un sistema que predice la evolución de

la infección en pacientes ingresados. Este sistema, combina el Big Data e IA (Inteligencia Artificial), procesando miles de datos –de indicadores tales como síntomas, comorbilidades o resultados de pruebas de laboratorio- para ofrecer un pronóstico personalizado de cada paciente. De esta manera, se puede saber con antelación cuál va a ser el futuro de la infección y cómo va a responder a la misma el cuerpo del paciente, y de esta manera predecir posibles patrones en la evolución de otros pacientes.

<https://campussanofi.es/e-professionals/noticias/big-data-covid19/#:~:text=El%20Big%20Data%3A%20un%20gran,y%20lograr%20acabar%20con%20%C3%A9l.>

CASTILLA-LA MANCHA Y BIG DATA

Big Data aplicado al ámbito de la salud – Atención Primaria

El Sescam utilizó millones de datos clínicos de pacientes para mejorar los diagnósticos de Atención Primaria. Es el único proyecto de Big Data dirigido a médicos de familia que se desarrolla actualmente en las comunidades autónomas. Tras un año implantado en la gerencia de Alcázar de San Juan, el proyecto Sapiens se va a extender a todos los centros de salud de la región a primeros de 2019. Es el primero de España que cuenta con esta herramienta, lo que demuestra su gran apuesta en materia de innovación tecnológica. El SESCAM incorporaba esta tecnología en 2019, dentro de su estrategia de innovación tecnológica, y permite a los profesionales acceder a actividad asistencial registrada desde enero de 2011, tanto en Atención Primaria como Hospitalaria.

14 millones y medio de datos clínicos de pacientes castellano manchegos están ya en una plataforma que ayuda a los médicos de familia a realizar diagnósticos con mayor información: la que proporcionan las historias clínicas generadas durante los últimos siete años por dos millones y medio de pacientes.

Este sistema inteligente se llama Sapiens. Analiza las historias clínicas y ayuda al profesional a diagnosticar y tratar determinadas patologías. Lee los datos, evalúa resultados, identifica relaciones previamente desconocidas entre patologías de distintos pacientes y conecta todas las fuentes generadas.

<https://sanidad.castillalamancha.es/saladeprensa/notas-de-prensa/el-sescam-comienza-extender-su-plataforma-de-ayuda-al-diagnostico-todos#:~:text=Sapiens%2C%20basado%20en%20la%20tecnolog%C3%ADa,las%20v%C3%ADas%20cl%C3%ADnicas%20del%20SESCAM.>

Big Data y la Transformación Digital

El Gobierno regional ha suscrito un protocolo de intenciones con la multinacional International Business Machines (IBM) para que se incorpore al centro de conocimiento sobre computación en la nube creado en el Centro Regional de Innovación Digital, un espacio orientado a desarrollar soluciones de administración digital que acerquen los servicios públicos a la ciudadanía, así como a impulsar la transformación digital del tejido empresarial castellano-manchego.

<https://www.ibm.com/blogs/think/es-es/2020/12/14/ibm-asesora-a-castilla-la-mancha-en-su-transformacion-digital/>

<https://objetivocastillalamancha.es/contenidos/firmado-protocolo-ibm-incorporacion-al-centro-regional-innovacion-digital>

Big Data en la Universidad de Castilla-La Mancha

Creación de un aula de Big Data y Ciberseguridad, que recibirá el nombre de Oesía-UCLM y realizará actividades como proyectos de investigación y desarrollo tecnológico y varios programas de formación como seminarios y becas y, entre otras, fomentar las vocaciones tecnológicas entre los estudiantes. Valdepeñas ya cuenta con un centro de ciberseguridad propio con alcance nacional. Collado ha reconocido estar «orgulloso y muy contento» por la vinculación con Oesía, lo que supondrá un «crecimiento» para la entidad.

https://www.abc.es/espana/castilla-la-mancha/abci-uclm-abre-centro-data-y-ciberseguridad-albacete-201907101436_noticia.html

II.XI. FASES DE UN PROYECTO BIG DATA E INDICADORES PARA LA DETECCIÓN DE NECESIDADES FORMATIVAS EN CASTILLA-LA MANCHA

Como ya se ha indicado desde el apartado de aproximación, el objetivo del presente proyecto es suplir una de las deficiencias del sistema de formación profesional para el empleo: la escasa anticipación a las tendencias del mercado laboral, con el consecuente desfase en la cualificación de los trabajadores/as de cara a los puestos de trabajo con mayor demanda. Por otro lado, el contexto actual, con la crisis sanitaria del COVID-19 que ha azotado la economía mundial, y más en territorios como España, dependientes de actividades terciarias, demanda soluciones que fomenten la recuperación económica. La formación es una palanca de cambio al potenciar el capital humano. Por ello, se tercia necesario la proyección de un sistema Big Data que integre los datos de las fuentes que den respuesta al objeto de estudio, los transforme y modele, de manera que sean válidos para extraer información de valor.

En el primer apartado, se ha detallado la evolución del Big Data desde la creación del término hasta la actualidad, además de una breve descripción de los elementos básicos que lo componen. Como se ha podido apreciar, este sistema está compuesto por multitud de procesos técnicos, de ahí la necesidad de un trabajo de campo preciso, donde sean los expertos/as quienes configuren el tipo de procesos, programas y características tecnológicas necesarias.

A continuación, tras el análisis documental realizado, se procede a desgranar los pasos a seguir en un proceso Big Data destinado al análisis de necesidades formativas, con herramientas y datos preestablecidos antes del análisis y conocimiento experto. Como se ha señalado, este punto se perfilará después como los perfiles informantes correspondientes.

En la infografía siguiente, se muestra los datos de acuerdo al objetivo del presente proyecto. Es una base, una relación de conceptos de la que partir, para que los investigadores, que no pertenecen al ámbito técnico del Big Data, puedan usar de guía para la aproximación con los expertos/as.

INFOGRAFÍA 1. Proceso Big Data: Detección necesidades formativas.



Fuente: Elaboración propia

II.XII.LIMITACIONES

No todos los anuncios de empleo publicados en línea están vinculados a una vacante real relacionado con una oferta de trabajo. A menudo, las agencias de empleo o las empresas colocan 'vacantes fantasmas' para probar el mercado, atraer solicitudes espontáneas e incluso crear listas de reserva de candidatos adecuados. Para ocupaciones con alta rotación, las empresas anunciarán permanentemente las vacantes para reemplazar empleados que se van. Identificar estos tipos de vacantes es difícil; sin embargo, la proporción de tales vacantes no distorsiona significativamente los datos y análisis de los requisitos del empleador.

El análisis de mercado tiene varias limitaciones:

- a) las vacantes en algunos sectores y ocupaciones están sobrerrepresentadas en los portales online. Hay que limpiar y depurar los datos.
- b) El uso de portales públicos difiere de los portales privados por las diferentes estructuras de empleo;
- c) Las habilidades/competencias/cualificaciones enumeradas en una oferta de empleo de vacante no reflejan el perfil de trabajo completo; empleadores tienden a enumerar solo las habilidades y calificaciones críticas para 'filtrar' a los solicitantes de empleo;
- d) Las ofertas de empleo deben ser legibles por máquina y utilizar un vocabulario y, dada la cantidad de datos (precisa de conocimiento técnico)
- e) La misma oferta de empleo puede publicarse en varios sitios web, necesidad de depuración

INFOGRAFÍA 2. Vida de una oferta de empleo



Fuente: Elaboración propia a partir de “CEDEFOP Online job vacancies and skills analysis”

II.XIII.EQUIPO DE PROYECTO

Para la creación de un sistema de detección de necesidades formativas a través de la tecnología Big Data, es necesario contar con un equipo de trabajo compuesto por diferentes perfiles técnicos que se complementen para lograr dar cobertura a cada una de las etapas del procesamiento de datos. Estos perfiles son orientativos, a posteriori se desarrollará el trabajo de campo para que dé respuesta a esta cuestión.

- *Director/a de proyectos que tenga experiencia en dirección y gestión de proyectos.*
- *Técnico/a analista*
- *Técnico/a de desarrollo. proyectos Big data*
- *Arquitecto/a de datos*
- *Consultor/a experto/a en diccionario web y vocabularios semánticos. Experiencia en entornos de desarrollo de ontologías y semántica web, NPL (Natural Language Processing) e IA (Inteligencia Artificial)*

II.XIV.INDICADORES

A acordar en el grupo experto junto con los técnicos correspondientes:

Problema: Detección de necesidades formativas entre los trabajadores/as de Castilla-La Mancha. Precisa:

- Qué sectores y/o actividades
- Territorio: Castilla-La Mancha
- Indicadores:
 - Ocupaciones recurrentes
 - Competencias/Habilidades/Cualificaciones
 - Formaciones, Cursos
- Herramientas

II.XV. FUENTES

Portales de empleo públicos

SEPECAM

INEM

Portales de empleo privados

INDEED

LINKEDIN

INFOJOBS

JOB TODAY

II.XVI.DATOS PRIMARIOS

Todos los datos que contienen las fuentes primarias como portales de empleo, servicios públicos de empleo, consejerías...

- Ocupaciones
- Vacantes de empleo
- Cursos formativos

Su obtención, en principio, será a través de alguno de las siguientes extensiones/formatos:

EXCEL

CSV

II.XVII.ALMACENAMIENTO

En un primer momento, tras el análisis de fuentes secundarias consultadas durante esta primera fase documental, se opta por idear la selección un almacenamiento **a través de servicios en la nube o cloud** por las siguientes ventajas que este sistema posee y que se han valorado de manera positiva frente a un almacenamiento clásico en un almacenamiento tradicional (dispositivos que estén instalados en un espacio de la organización en el que se guarden)

Mantenimiento

Los costes de un servicio en nube son menores que tener el propio hardware, ya que el mantenimiento de los servidores es por cuenta de la empresa a la que se compra el espacio. Esto supone un ahorro en dinero y tiempo.

Accesibilidad

Cuando se almacenan los datos en la nube, se puede acceder en todo momento, y desde cualquier punto con acceso a red.

Optimizar el espacio

Un servidor local ocupa espacio físico, mientras que el almacenamiento de datos en nube, optimiza el uso de los espacios.

Seguridad

Las empresas con servicios de *cloud* tienen una seguridad de la información muy elevada, lo que minimiza los ciberataques y los riesgos de pérdida de datos.

Reducción de costes

Cuando tu empresa invierte en servidores de almacenamiento, debe afrontar todos los gastos derivados. Si pagas por el almacenamiento, todos estos gastos secundarios van incluidos.

II.XVIII.TRANSFORMACIÓN

La transformación de datos es el proceso de convertir datos o información de un formato a otro. Este proceso se realiza para hacer convertir los datos a otra extensión de manera que sea posible tratarlos, procesarlos e interpretarlos en otro sistema en el que se puedan modelar y extraer información.

La transformación de datos implica el uso de un programa especial que es capaz de leer el lenguaje base de los datos brutos, determinar el formato en el que los datos que se deben procesar para que sea utilizable por el nuevo programa.

Gama de actividades dentro de la transformación:

- Limpiar
- Eliminar datos erróneos o nulos
- Eliminar duplicado de datos
- Determinar estructura

Scripting

Herramientas de ETL

Nube

Local

II.XIX.VISUALIZAR

De forma que la modelación de los datos sea comprensible y la comunicación sea concreta para dar presentación y entender entre usuarios y clientes.



Fuente: Imagen extraída del portal Freepik

II.XX. HERRAMIENTAS QUE PUDEN SERVIR PARA EL PROCESAMIENTO DE DATOS

Aunque este apartado se investigará en la fase de trabajo de campo, cuando se cuente con la experiencia y conocimiento de los experto/as consultados/as, se ha llevado a cabo una relación de las herramientas que sería interesantes presentarles por su presencia y reconocimiento entre las páginas expertas consultadas. Cubriendo todos los pasos de la misma: almacenamiento, procesamiento, análisis y visualización.

1. Hadoop



2. Elasticsearch



3. Apache Spark



5. Apache Kafka



6. Jupyter



7. Keras



8. Kibana



9. R Shiny



III. CONCLUSIONES

La escucha activa del mercado de trabajo en la región de Castilla-La Mancha es imprescindible para la permanente actualización de las políticas activas de empleo. Asimismo, en el contexto actual, con la transformación digital en permanente crecimiento y el avance de la digitalización en todos los ámbitos, hace necesario el replanteamiento de las cualificaciones y los cursos que forman al conjunto de personas trabajadoras de la región.

A través de este proyecto, se propone la creación una hoja de ruta para la configuración de un sistema Big Data que logre recolectar datos de diferentes fuentes de información (portales públicos, y privados), para detectar patrones de comportamiento del mercado de trabajo y de las competencias ligadas a él, de forma "regular y dinámica" y de esta forma se ajusten las formaciones y competencias a los requisitos del mercado y sus trabajadores/as.

El análisis documental de esta primera fase concluye con diversas propuestas de herramientas y pasos a seguir. Será en la fase de campo, gracias al conocimiento y experiencia de los informantes clave que se seleccionen, cuando se recojan los detalles técnicos correspondientes a la operativización de un proceso de datos como el que se pretende. Una de las conclusiones a las que se ha llegado, tras la revisión es la importancia de la combinación de fuentes, de donde surge un plus de valor. Los portales de empleo tienen un gran potencial, pero no son los únicos que contienen información sobre las demandas del mercado laboral, también los servicios públicos de empleo, y las consejerías pueden ofrecer datos que sirvan para su cruce.

Cada vez existen más herramientas de procesamiento de datos (depuración, análisis, visualización, transformación...), así como más opciones de almacenamiento. Esto plantea un horizonte positivo para la creación de este tipo de sistemas para entidades de mayor y menor tamaño, así públicas como privadas. Como ejemplo, el Servicio de Salud de Castilla-La Mancha (SESCAM) es pionero en integrar el sistema Big Data en conjunto con la Inteligencia Artificial. Gracias a este proyecto, se pretende acelerar el diagnóstico y la detección temprana de problemas.

Se han detectado ciertas limitaciones de un sistema Big Data que hay que tener en cuenta en el momento de elaboración metodológica de la recogida de información, para que sean planteadas de manera rigurosa a los expertos:

- Leyes de protección de datos. Acceso a fuentes públicas y privadas. Tipos de convenios posibles para el la recogida de datos (se tiene conocimiento que existen entre plataformas como Infojobs y el Inaem de Aragón)
- Funcionamiento preciso de los sistemas ontológicos y semánticos, así como de las posibilidades de Scrapping de redes sociales, o portales como Linkedn.
- Recursos Humanos; Perfiles competenciales necesarios para la puesta en marcha de un proyecto estas características.
- Costes

IV. ACTIVIDADES REALIZADAS

OBJETIVOS PREVISTOS EN EL DESARROLLO DE LA FASE	GRADO DE CUMPLIMIENTO
Analizar datos de fuentes secundarias, informes, estadísticas, estudios actuales que permitan contextualizar la tecnología Big Data a nivel nacional y regional, e identificar iniciativas existentes de éxito	Completo
Recopilar datos sobre las características de cada una de las fases necesarias para implantación de un proyecto que requiera de la tecnología Big Data	Completo
Revisar bibliografía sobre indicadores y/o variables que ayuden a la detección de necesidades formativas	Completo

ACTIVIDADES REALIZADAS	FECHA INICIO	FECHA FIN
Selección de estudios, publicaciones, datos, estadísticas actuales disponibles sobre el desarrollo de un proyecto Big Data, que permita concretar de manera clara las fases a completar	04/01/2021	31/03/2021
Contextualización la situación del Big Data, su evolución y tendencias, así como casos de éxito reconocidos	04/01/2021	31/03/2021
Búsqueda de información que permita el posterior planteamiento sobre los indicadores y/o variables más oportunas para la detección de necesidades formativas	04/01/2021	31/03/2021

FASE 2. TRABAJO DE CAMPO

INTRODUCCIÓN

La fase de trabajo dentro de una investigación pone el foco en la recogida de información a través de la aplicación de técnicas metodológicas que se ajusten a los objetivos del proyecto en concreto. La información recabada en este caso, tiene como fin definir los puntos para la integración de la tecnología Big Data en la FPE y crear una plataforma que de servicio y conocimiento sobre las necesidades de las personas trabajadoras de Castilla-La Mancha, y de esta forma sea posible ajustar las demandas actuales del mercado con la formación ofertada por el sistema de formación para el empleo.

En esta fase se selecciona la metodología que mejor se ajusta a las postulaciones del estudio, las técnicas adecuadas dentro de ella y se programa la ejecución del trabajo de campo. Para el presente proyecto el equipo técnico se decanta por la metodología cualitativa, al poseer datos anteriores y necesitar un conocimiento experto en primera persona para completar los postulados previos dispuestos en la memoria entregada a la junta del proyecto PE/2019/02.

En los próximos capítulos se profundiza en las particularidades de este trabajo de campo, en cuestiones como la delimitación de la población experta a entrevistar; las estrategias empleadas para acceder a la muestra; la justificación, de la técnica y las herramientas empleadas en la recopilación la información; aspectos clave para un correcto desarrollo del trabajo de campo. Finalmente se aborda el cronograma de campo.

I. ENTREVISTAS EN PROFUNDIDAD

I.1. Justificación de la técnica

Las entrevistas en profundidad son procesos comunicativos que permiten la extracción de información por parte de un/a investigador/a (Alonso 1994, en Valles 1997). A partir de la aplicación de esta técnica cualitativa, el investigador de manera flexible obtiene una gran riqueza informativa en las palabras de los/las profesionales entrevistados/as, además, su naturaleza abierta, le permite indagar en profundidad sobre ideas, valoraciones, experiencias y opiniones de cualquier cuestión temática.

Las ventajas principales de la aplicación de entrevistas en profundidad en un proyecto son:

El estilo especialmente abierto de esta técnica permite la obtención de una gran riqueza informativa (intensiva, de carácter holístico o contextualizada), en las palabras y enfoques de los entrevistados.

Proporciona al investigador la oportunidad de clarificación y seguimiento de preguntas y respuestas (incluso por derroteros no previstos), en un marco de

interacción más directa, personalizado, flexible y espontáneo que la entrevista estructurada o de encuesta

Proporciona al investigador la oportunidad de clarificación y seguimiento de preguntas y respuestas (incluso por derroteros no previstos), en un marco de interacción más directa, personalizado, flexible y espontáneo que la entrevista estructurada o de encuesta

Frente a las técnicas cualitativas de observación, la entrevista en profundidad es más capaz y eficaz en el acceso a la información difícil de obtener sin la mediación del entrevistador o de un contexto grupal de interacción.

Fuente: Elaboración propia a partir de Valle, 1997

La razón principal para la elección de esta técnica radica en la importancia de contar en primera persona con el discurso y conocimiento del informante clave. Se necesita que el investigador escuche y cuestione las dudas principales con una técnica flexible que le permita ahondar en puntos de interés para el desarrollo de la investigación.

II.II. Selección de informantes clave

Una parte sustancial de todo diseño de investigación es la selección de la población a través de la cual se extrae la información en relación con el objeto de estudio.

Como el objetivo de las entrevistas es la definición de los ejes temáticos en torno a la aplicación de la tecnología de Big Data como posible medio precursor de detección de necesidades formativas, es preciso realizar una adecuada selección de los informantes clave que serán entrevistados. Para contar con una perspectiva veraz y precisa, fueron seleccionados los siguientes perfiles de expertos/as:

- ✓ Profesionales del ámbito de las Tecnologías de la Información y la Comunicación
- ✓ Profesionales del ámbito de la tecnología Big Data.

Cuadro 1. Perfil experto y motivo de selección

PERFIL EXPERTO	MOTIVO DE SELECCIÓN
Expertos en Transformación Digital (TIC) y tecnología Big Data	- Proporcionan información (asesoramiento) en materia técnica para conocer previamente qué es la tecnología Big Data y en qué consiste la analítica de datos. - Proporcionan conocimiento enfocado a proyectos desarrollados con tecnología Big Data. - Nos ofrecerán soporte experto para elaborar una memoria técnica para el desarrollo de una plataforma de Big Data

- Además, nos proporcionarán las primeras pinceladas sobre los mejores procedimientos analíticos a llevar a cabo en la región para conseguir mejoras en la detección de necesidades y en la mejora del empleo en la región.

Fuente: Elaboración propia

Dado que este es un proyecto derivado de una investigación anterior, *PE/2019/02 “Asesoramiento a pymes sobre la aplicación de la tecnología Big Data en la formación de sus trabajadores y la creación de empleo”* la captación de los perfiles clave para la ejecución de las entrevistas se ha desarrollado de una manera sencilla, al contar con una BBDD ya estructurada con los detalles de los informantes clave.

Por lo que los pasos que se siguieron fueron:

- ✓ Preparación de emails explicando el proyecto y pidiendo colaboración en el mismo a través de entrevistas telefónicas o telemáticas.
- ✓ Contactación y seguimiento a través de llamadas e emails de los participantes: puesta en contacto y contextualización del seguimiento del proyecto llevado a cabo el año pasado PE/2019/19

Cuadro 2. Profesionales entrevistados, clasificados por perfil informante, tipo y fecha de entrevista

Perfil informante	Profesional entrevistado/a ¹	Tipo de entrevista	Fecha de entrevista
Expertos en Transformación Digital (TIC) y Big Data	Miguel Ángel: Ingeniero Superior en Informática e Ingeniero Técnico en Informática de Gestión.	Telefónica	21 de mayo de 2021
Expertos en Transformación Digital (TIC) y Big Data	Octavio: Consultor experto en tecnologías de la información, incluyendo la tecnología Big Data	Telefónica	25 de mayo de 2021
Experto en análisis de datos y trabajador Alto Analytics	Vishrut: Data Analyst Alto Data Analytics	Telefónica	27 de mayo de 2022

¹ Se evitan el detalle de los apellidos para garantizar la anonimización de la información

Todas las entrevistas se graban en audio para el posterior análisis del discurso. Gracias a este método es posible asegurar la información para no perder detalle y veracidad de la información recabada. Además, en este tipo de análisis, se necesitan las transcripciones *verbatim*s (frases literales del informante clave) que sirven de apoyo tanto para el investigador, como para reforzar las premisas resultantes de la investigación.

I.III. Diseño del guion de entrevista

Previamente, a la ejecución de las entrevistas cualitativas, se hace imprescindible elaborar un guion con las cuestiones de interés que se abordarán durante la misma.

La entrevista en profundidad es una técnica flexible y abierta que permite redirigir el discurso en función de la información que vaya emergiendo, no obstante, elaborar previamente un guion permite controlar las variables (preguntas) de interés que sean imprescindibles abarcar, de esta forma, se garantiza el cumplimiento mínimo de los objetivos generales y específicos del estudio.

Para la redacción de este guion, se tiene presente en todo momento la información que se quiere recoger para responder adecuadamente a los objetivos planteados en el estudio.

Cuadro 3. Guía para las entrevistas a expertos/as en tecnología Big Data y análisis de datos

Bloques	Objetivos de información
Bloque I. Introducción a la tecnología Big Data	- Contextualización del objeto de estudio. Proporcionan información para conocer el diseño de una plataforma que aloje tecnología Big Data - Identificación de los procedimientos analíticos y arquitecturas digitales para el tratamiento de datos a gran escala
Bloque II. Big Data, foco a la detección de necesidades formativas	- Información técnica sobre el procedimiento para abordar la temática seleccionada (detección de necesidades formativas en la región de Castilla-La Mancha). Pasos a seguir y equipo de trabajo. - Información que de soporte para enfocar correctamente la memoria técnica para el desarrollo y creación de una plataforma.

II. GRUPO DE TRABAJO EXPERTO

II.I. Justificación de la técnica

La selección de ejecutar un grupo de trabajo experto como técnica de investigación cualitativa se fundamenta en la necesidad de recogida de datos y conocimiento a través de la interacción de las perspectivas y propuestas de ejecución que posean los profesionales y/o expertos/as en materia tecnológica y Big Data, o informantes clave con experiencia en la planificación y ejecución de la formación profesional, y ejecución de proyectos.

Para ello se plantea formar un grupo experto de entre 3 y 6 participantes, para favorecer el intercambio de comunicación, conocimiento y la colaboración entre los participantes. La duración de la mesa de trabajo será de mínimo 60 minutos, sin tener un máximo temporal, entendiéndose que no excederá los 180 minutos.

El objeto de este grupo de trabajo experto es compartir experiencias, propuestas e información sobre el diseño de un proceso de detección de necesidades formativas a través de la tecnología Big Data. Los perfiles están relacionados con con el análisis de datos, la puesta en marcha y ejecución de proyectos tecnológicos, y se incluye un perfil que corresponde con una persona del equipo técnico encargada de aportar el conocimiento y *expertise* de la formación para el empleo.

Dadas las circunstancias contextuales derivadas de la crisis sanitaria del COVID-19, se decide que la mejor forma de interactuar con los integrantes es a través de una plataforma *online*. La elegida es s a través de *Google Meet*. Dada su sencillez, gratuidad y seguridad

II.II. Briefing

Para realizar un correcto desarrollo de la reunión de trabajo es imprescindible la elaboración previa de una guía, que en este caso se denominará *Briefing*, que recoja las principales áreas temáticas a abordar durante el grupo

Las reuniones técnicas tienen como objetivo general el intercambio de conocimientos, presentación y discusión de experiencias, suministrando un escenario para el análisis de temas relevantes y actuales de interés común, en este caso, la tecnología Big Data

La dinámica será similar a la de un grupo de trabajo (el grupo de discusión tiene menos intervención por parte del moderador). De modo que el/la técnico/a correspondiente pueda mostrar también aportaciones, sugerir ideas, cambios... que se convierta en un lugar de intercambio de ideas a través de la contraposición de enfoques y experiencias.

- i. Valoración /diagnóstico de situación: Puesta en antecedentes del proyecto
- ii. Identificación de puntos, fases de desarrollo de un proyecto de estas características
- iii. Identificación herramientas y estructuras necesarias

- iv. Temporalización y asignación de perfiles para el equipo de trabajo
- v. Análisis de riesgos, costes

Al final de la fase se anexa el documento utilizado.

II.III. Participantes

Perfil Profesional	Informante Clave*
Senior Brand & Service Designer at The Cocktail / Wunderman Thompson Consulting	Enrique
Certification Manager Hydromechanical Systems at Airbus/Project management and systems integration	Roberto
Consultant at Business Intelligence/Data Analyst	Vishrut

* Los apellidos están en la BBDD, se ocultan en el informe por anonimizar y no infringir la ley de protección de datos en su publicación.

III. GESTIÓN Y PLANIFICACIÓN DEL TRABAJO DE CAMPO

- Documentación a través del proyecto PE/2019/02: 5 días ■
- Diseño de las herramientas de campo (ambos guiones): 5 días ■
- Subida de las herramientas de campo al Registro Electrónico: 1 día ■
- Revisión BBDD y contactación entrevista y grupo de discusión: 4 días ■
- Realización de entrevistas: 3 días ■
- Realización de grupo de trabajo: 1 días ■

ABRIL 2021						
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18

19	20	21	22	23	24	25
26	27	28	29	30		

MAYO 2021						
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

IV. ACTIVIDADES REALIZADAS

OBJETIVOS PREVISTOS EN EL DESARROLLO DE LA FASE	GRADO DE CUMPLIMIENTO
Diseñar y realizar el trabajo de campo para la recogida de la información cualitativa a través de entrevistas en profundidad a expertos en tecnología Big Data, análisis de datos y/o necesidades formativas de la región de Castilla-La Mancha, que respondan a los objetivos generales y específicos del proyecto, y a través de un grupo de trabajo que ponga en común experiencias e ideas sobre el desarrollo de la aplicación.	Completo
Identificar los requerimientos y tareas a realizar en cada una de las fases de un proyecto de análisis de datos a través de la tecnología Big Data	Completo
Identificar los perfiles profesionales, las fuentes de información, así como herramientas requeridas en un proceso de detección de necesidades formativas de los trabajadores/as	Completo

ACTIVIDADES REALIZADAS	FECHA INICIO	FECHA FIN
Se profundizará en el diseño metodológico del trabajo de campo cualitativo, así como en la adecuación de los perfiles de los informantes clave para la consecución de los objetivos previstos	01/04/2021	31/05/2021

Diseño de guiones temáticos de entrevistas según el objetivo que sea previsto alcanzar	01/04/2021	31/05/2021
Elaboración de un <i>briefing</i> para el funcionamiento del equipo de trabajo	01/04/2021	31/05/2021
Creación de una base de datos con informantes clave que puedan participar en las entrevistas cualitativas	01/04/2021	31/05/2021
Creación de una base de datos con informantes clave que puedan participar en el grupo experto	01/04/2021	31/05/2021
Selección y contacto con los participantes de las entrevistas	01/04/2021	31/05/2021
Selección y contacto con los participantes del grupo de trabajo	01/04/2021	31/05/2021
Gestión y planificaciones del trabajo de campo	01/04/2021	31/05/2021
Grabaciones en formato audio de las entrevistas cualitativas y el grupo de trabajo	01/04/2021	31/05/2021
Gestión de la documentación para la justificación de la ejecución de las técnicas	01/04/2021	31/05/2021

FASE 3. ANÁLISIS

INTRODUCCIÓN

En la presente fase se lleva a cabo el análisis cualitativo de la información recopilada durante la fase de trabajo de campo para la redacción de resultados. Este análisis se realiza tanto de las entrevistas cualitativas realizadas a los expertos en *big data* rescatados del anterior proyecto PE/2019/02, como del grupo de trabajo creado para determinar la forma óptima de definir las pautas de un proyecto de estas características.

El objetivo general del estudio es definir las líneas principales de un sistema de analítica de datos que permita la detección de necesidades formativas de la región. A su vez, este objetivo se concreta en los siguientes objetivos específicos:

- Definir cada una de las fases y puntos clave a ejecutar para el correcto desarrollo del proyecto
- Identificar los KPI's
- Determinar los perfiles profesionales y/o expertos/as adecuados para la posterior ejecución del proyecto
- Determinar y evaluar las fuentes de las que se puede recoger datos

El proceso analítico contará con varias etapas ajustándose a la metodología de recogida de información empleada:

- **Análisis de las entrevistas cualitativas:** Por medio de las entrevistas cualitativas se pretende esclarecer y definir el conjunto de tareas y pasos a seguir en un proyecto de análisis de datos como este, para la detección de competencias demandadas en el mercado, así como adelanto de tendencias de mercado por medio de las ofertas de trabajo.

Las entrevistas, tal como se detalló en la fase anterior, se desarrollaron de manera abierta, con un guion semiestructurado, sin categorías y valoraciones preestablecidas, de forma que los distintos expertos y expertas expresaran con libertad sus opiniones y experiencias sobre el objeto de estudio, en este caso, la tecnología *big data*. Una vez realizadas y grabadas vía audio, con el correspondiente permiso del participante, se procede a la transcripción literal de las entrevistas, reparando sobre todo en aquellas partes cuya información respondiese a los objetivos del estudio y enriqueciera el proyecto con información adicional de interés. Las transcripciones literales facilitaron el siguiente eslabón en el proceso analítico: el análisis de contenido. Esta modalidad analítica engloba un conjunto de técnicas que buscan sistematizar el contenido de mensajes, en este caso procedentes de textos transcritos, con el objetivo de efectuar deducciones lógicas. Para ello, se emplea la “categorización”, esto es: “una operación de clasificación de elementos

constitutivos de un conjunto por diferenciación, tras la agrupación por analogía, a partir de criterios previamente definidos” (Bardin, 1996). Se extraen de estas *verbatim*s: citas literales de los entrevistados que reflejan y apoyan las ideas y conclusiones extraídas del discurso.

- **Análisis del grupo experto:** Un análisis de las intervenciones y contenidos que fueron tratados durante la reunión virtual, por medio del análisis de contenidos de las intervenciones y otros aspectos que sean reseñables de cara a cumplir con los objetivos del estudio. Para el análisis se hace uso de las grabaciones de audio, de igual manera que con las entrevistas, por la practicidad y el fácil acceso a estas.

Todo ello se agrupará en un análisis de los puntos a tener en cuenta para el desarrollo técnico de una plataforma que englobe las características y objetivos del proyecto.

I. ANÁLISIS DE LOS RESULTADOS

La información analizada en las entrevistas y grupo de expertos que se presenta en las próximas líneas ha sido clasificada en ejes temáticos. Se establecen y delimitan estos ejes para reunir de manera estratégica la información más precisa y determinante sobre el problema y objetivo de investigación.

¿Qué aporta la tecnología big data a un proyecto como el que se postula sobre detección de necesidades formativas a través de la tecnología big data?

Este punto pretende no solo comprender el término, sino también la experiencia del entrevistado con este tipo de tecnología para lograr un acercamiento a él. Según los puntos en común de los discursos de los expertos/as, el *big data* es un conjunto de arquitecturas tecnológicas que permiten almacenar y manejar datos de naturaleza “no estructurada” en grandes cantidades, para abastecer necesidades de negocio y ofrecer soluciones a problemas detectados. Por ello, afirman que el *big data* va más allá de la analítica de datos más tradicional, y se centra en tratar y estudiar aquella información más difícil de explotar porque no dispone de un formato específico como es el caso de las imágenes, los vídeos, los discursos, etcétera, pero que también aportan valor.

Recuerdan y dan valor al punto más importante del proyecto, las preguntas clave y su “ecosistema de proyecto”, es decir, las prescripciones que se plantean en un comienzo para llevar a cabo las cargas de trabajo. La detección de necesidades formativas es un proyecto que puede emplear la tecnología *big data* y los expertos lo perciben perfectamente viable y de una complejidad baja, por la estructura de los datos que se requieren recoger. Aun así, destacan la importancia de una prescripción técnica y de recursos humanos óptima para el desarrollo del proyecto de una manera eficaz.

“Un proyecto de dato...no es un proyecto de desarrollo de software...como todo tiene una interfaz digital(...) lo que pasa que eso es lo de menos en un proyecto de dato. Lo que menos os tiene que preocupar. (...) en un equipo de desarrollo de datos no hay

“para crear una estructura de data análisis se necesita un CTO (...) esa persona crea una infraestructura que rescata los datos que se requieren recoger y analizar a posteriori...esa es la base de crear una herramienta y un análisis” (E_3)

ocho programadores, hay dos o tres, como mucho, pero no más. Porque si no el proyecto no sale” (E_1)

Para los expertos consultados, el *big data* es un proceso que genera valor y oportunidades para aquellos proyectos que son adecuados para el uso de su tecnología. La importancia de este recurso tecnológico va asociado a la puesta en valor que hace de los datos y la información.

“Estamos hablando de big data por una única cuestión, porque queremos sacar el valor de los datos. No es una cuestión de volumen, no es una cuestión de variedad, podemos estar sacándolo en datos de 1000 empleados cómo lo podemos estar sacando en datos de un millón. Dependiendo de cómo sean podremos sacar mejores o peores patrones. Lo que quiere el ser humano es ser capaz de saber lo que va a pasar mañana, por eso antes del Big Data llevábamos años y años con algoritmos, con investigación y proyectos de data mining, es lo mismo, pero le hemos ido cambiando el nombre. Hoy ya no se le llama tanto Big Data, es un término que empieza a estar en decadencia, empezamos a hablar más de Data Science, llámalo como tú quieras. La tecnología ha permitido coger datos de todos los procesos, podrías analizar esos datos para sacar información de por qué están pasando las cosas o qué va a pasar mañana basándome en lo que pasó ayer” (E_3)

Fases de un proyecto de analítica de datos: Big Data

Definición de las necesidades y las metas:

La tecnología *big data* responde perfectamente a las necesidades de un proyecto de las características del presente. Los expertos, al abordar la temática del mismo, confirman su viabilidad. El valor que presenta una herramienta así, es la capacidad de orientar las políticas de empleo, así como los currículos de las formaciones, gracias a la capacidad que tiene la funcionalidad de la herramienta para analizar en profundidad el mercado de la región de Castilla-La Mancha. Las fuentes de datos de donde extraer la información son accesibles y su contenido está bastante estructurado.

“Un proyecto no arranca porque tengamos datos, arranca porque tenemos un problema, tú tienes que definir el problema qué quieres resolver y tienes que establecer tus metas de negocio. ¿Qué necesito? El perfil de las personas que van a trabajar en no sé qué, o sea tú defines tus metas de negocio y tus KPIS, cuáles son los indicadores de éxito de ese proyecto. El problema es que muchos proyectos arrancan porque hay datos y nadie ha definido unos KPIS” (E_1)

Lo que se quiere recoger y analizar a través de la herramienta:

- Competencias laborales más demandadas en el mercado laboral
- Ocupaciones con más (y menos) ofertas de empleo
- Ofertas según espacio de tiempo
- Análisis según zona geográfica
- Tendencias
- Cursos más ofertados
- Relación entre formación y ofertas de empleo, correspondencia

Establecer un equipo de trabajo con expertos en Big Data:

El siguiente paso consiste en establecer un equipo de expertos y expertas para que den soporte profesional. Tras la escucha de los expertos, se entiende que un equipo debe contar, mínimo, con los siguientes perfiles:

- Dirección técnica del proyecto con experiencia en gestión y coordinación de proyectos con tecnología big data, y conocimiento en dirección
- Analistas con experiencia en proyectos de esta tipología
- Arquitecto de datos que construya la herramienta y su interfaz
- Consultor de estrategia de proyecto, que plantee y marque los KPI's a desarrollar en la puesta en marcha de la herramienta

“Un grupo de 10 personas para esto sobra... necesitaréis tres o cuatro perfiles que trabajasen con datos(...)...un par de perfiles de desarrollo software...estos más básicos, luego dos personas de un perfil superior, uno un perfil más arquitecto (...) y una persona de matemáticas, es decir, que te haga modelos matemáticos que les den respuesta a los KPI's planteados” (E_3)

“Necesitáis un consultor, incluso una pequeña consultoría, para ver cuál es la solución técnica que os sirva” (E_2)

Determinar los datos a integrar:

En esta fase del proyecto de Big Data, con ayuda del equipo técnico, habrá que decidir qué fuentes de datos se van a utilizar para resolver las necesidades y los KPIS establecidos. Se sugieren dos tipos de fuentes de datos que nutran a la plataforma.

- Oferta de empleo:
 - Sepecam
 - Indeed
 - Infojobs
 - LinkedIn
 - RRSS
- ...
- Formación para el Empleo:
 - Fundae
 - Sepe
 - Consejería de Empleo de Castilla-La Mancha (oferta de formaciones de todas las modalidades)

Desarrollo

- KPI'S, conocimiento en profundidad de las preguntas que se quieren responder. Objetivo principal: Responder a las necesidades de cualificación a través de una plataforma que extraiga la información del mercado y sus requerimientos.
- Creación de arquitectura de data que capture los datos de las fuentes de información. Plataforma donde se visualicen los datos y pueda ser accesible a miembros de la administración.
- Calidad del dato: La depuración y preparación de los datos, según los expertos y expertas, es la tarea que más tiempo consume dentro del proyecto. Concretamente en esta fase, se depuran, tratan y preparan los datos adquiridos para empezar a trabajar. Son los perfiles de analistas los que llevarían a cabo este proceso. Durante la depuración (identificación de la naturaleza del dato, contemplación de su formato, homogeneización de datos de distintas fuentes, eliminación de datos nulos, etc.) Se detecta la calidad de los mismos y se explora si ofrecerán lo necesario para seguir avanzando en los siguientes pasos.

Establecer la inteligencia artificial (data mining, big data, data science, etc.) más adecuada

La aplicación de la inteligencia artificial es muy variada y cada una de las técnicas está orientada a satisfacer necesidades muy distintas. Los expertos mencionan infinidad de conceptos, el perfil técnico (matemático) será el encargado con los KPI's establecidos al comienzo del desarrollo de la plataforma en seleccionar y crear la fórmula más adecuada para extraer la información demandada, así como predecir tendencias futuras.

Otras consideraciones en un proyecto de Big Data:

- **Coste:** según los expertos/as es importante un buen asesoramiento de profesionales que analicen de forma adecuada las necesidades de la entidad para establecer un trabajo analítico adecuado y conciso que no genere costes adicionales. Los proyectos pueden ir desde los 10.000 euros, hasta la cifra que el cliente esté dispuesto a pagar. Ha tener en cuenta que al empezar un proyecto desde cero, los perfiles con mayor cualificación requieren un mayor coste de contratación por las competencias tan especializadas que poseen.
- **Protección de datos:** es importante que la anonimización de datos sea estricta. Acuerdos con las fuentes de información (desde Infojobs hasta la Consejería de Empleo) basados en este principio.

Las posibilidades del Big Data en la mejora del empleo y la formación de los trabajadores/as

Como indican los expertos, es imprescindible definir *a priori* cuál es la necesidad a la que se quiere dar respuesta y cuáles son las metas a cubrir (KPI,s) Un consultor experto es el perfil encargado de esta tarea. En este caso la meta general, siguiendo los objetivos del proyecto, es la anticipación de las tendencias del mercado y las necesidades formativas de los/las trabajadores/as para favorecer la empleabilidad en la región, lo que permitiría ajustar las políticas de empleo y la oferta de formaciones (actualizar currículos y lanzar las más demandadas por el mercado laboral).

Los expertos recomiendan focalizar y definir dentro de esta necesidad general qué indicadores o KPIS concretos son necesarios abordar para lograr dar respuesta a la misma. En este punto cabría preguntarse, ¿cómo nos anticipamos a las tendencias del mercado? ¿cómo nos anticiparíamos a conocer las necesidades formativas? ¿Qué indicadores influyen en la configuración de las tendencias de mercado y en las necesidades formativas? Una vez definidas estas necesidades o metas, se procedería a la búsqueda de los datos y las fuentes adecuados para cubrir estas necesidades.

Abordar la sincronía entre la demanda de las empresas y la oferta formativa en la región, es una necesidad pues se observa cierta falta de sincronía entre las necesidades de las empresas y la formación con la que llegan los alumnos y alumnas para realizar las prácticas en su centro de trabajo.

En la siguiente fase, de difusión y elaboración de documentos divulgativos, se diseñarán los documentos que contengan la información técnica y divulgativa de los resultados obtenidos de manera detallada para su consulta posterior.

II. CONCLUSIONES

Para cumplir con los objetivos del proyecto es necesario que se cuente con unos KPI's claros, derivados de una consultoría previa, ya sea interna con un consultor propio, como externa.

KPI's planteados:

- Competencias laborales más demandadas en el mercado laboral
- Ocupaciones con más (y menos) ofertas de empleo
- Ofertas según espacio de tiempo
- Análisis según zona geográfica
- Tendencias
- Cursos más ofertados
- Relación entre formación y ofertas de empleo, correspondencia

Equipo de trabajo:

- Dirección técnica del proyecto con experiencia en gestión y coordinación de proyectos con tecnología big data, y conocimiento en

- Analistas con experiencia en proyectos de esta tipología
- Arquitecto de datos que construya la herramienta y su interfaz
- Consultor de estrategia de proyecto, que plantee y marque los KPI's a desarrollar en la puesta en marcha de la herramienta

Las fuentes de datos sugeridas son:

- Ofertas de empleo:
 - Sepecam
 - Indeed
 - Infojobs
 - LinkedIn
 - RRSS
- ...
- Formación para el Empleo:
 - Fundae
 - Sepe
 - Consejería de Empleo de Castilla-La Mancha (oferta de formaciones de todas las modalidades)
-

Otras consideraciones del proyecto son: costes y protección de datos.

III. ACTIVIDADES REALIZADAS

OBJETIVOS PREVISTOS EN EL DESARROLLO DE LA FASE	GRADO DE CUMPLIMIENTO
Analizar la información obtenida a través de las técnicas cualitativas de investigación llevadas a cabo durante el trabajo de campo	Completo
Responder a los objetivos generales y específicos de la investigación mediante el tratamiento en conjunto de los datos	Completo
Detallar las conclusiones extraídas de tal manera que sirvan de base para la elaboración y diseño de los productos finales	Completo

ACTIVIDADES REALIZADAS	FECHA INICIO	FECHA FIN
Escucha de los audios de las entrevistas cualitativas de los diferentes perfiles seleccionados, así como del grupo de trabajo	01/06/2021	30/07/2021
Selección de verbatims más representativos toma de notas sobre la información más relevante a tener en cuenta para alcanzar los objetivos previstos	01/06/2021	30/07/2021

Análisis de contenido de todas las entrevistas cualitativas y el grupo de trabajo acorde con los objetivos específicos del estudio	01/06/2021	30/07/2021
Identificación de las fases a seguir para la correcta implantación de un sistema de análisis de datos a través de la tecnología Big Data, así como de las fuentes de información y variables a seleccionar	01/06/2021	30/07/2021
Redacción del informe de la fase con el análisis en profundidad del trabajo de campo	01/06/2021	30/07/2021

FASE 4. EDICIÓN, PUBLICACIÓN Y DIFUSIÓN

INTRODUCCIÓN

El presente informe equivale a la *Fase 4* del proyecto “*Prospección y detección de necesidades formativas a través de la tecnología Big Data en Castilla-La Mancha*”

En los próximos apartados se presentan las actividades desarrolladas durante el proceso para llevar a cabo los objetivos de la presente fase:

- Diseñar las infografías en formato digital dedicadas a la presentación del uso de la tecnología Big Data y su contextualización a nivel nacional y regional
- Elaborar productos que definan de manera precisa los pasos a seguir para el correcto desarrollo de un proyecto de analítica de datos a través de la tecnología Big Data
- Redactar informe ejecutivo con las conclusiones principales alcanzadas durante el estudio
- Difusión de contenidos

1. INFOGRAFÍAS CON LAS CONCLUSIONES DEL ESTUDIO

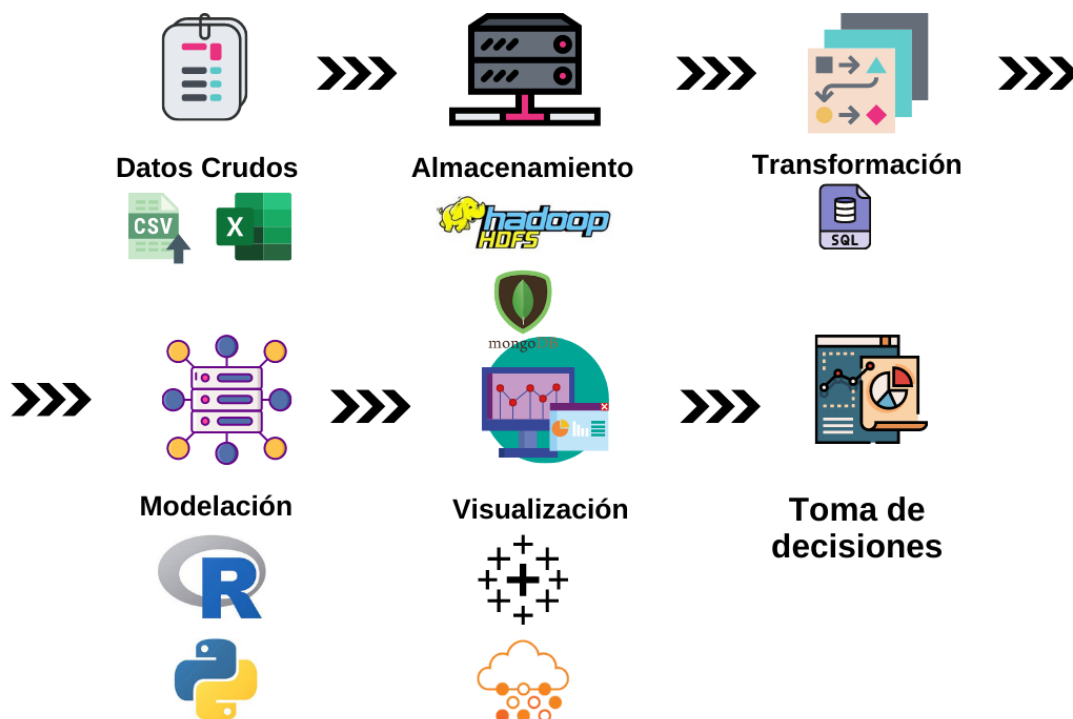
En este capítulo se presentan las infografías elaboradas para su consulta y difusión con las conclusiones del proyecto.

La primera infografía denominada “*Fases y herramientas de un proyecto Big Data*” donde se exponen tanto los pasos a seguir en la explotación y análisis de datos, como las herramientas clave a utilizar.

La segunda infografía denominada “*Uso de la tecnología Big Data a nivel nacional y regional*”, presenta información sobre el uso que presenta esta tecnología en el territorio a través de ejemplos

1. Infografía nº1: “Fases y herramientas de un proyecto Big Data”.

FASES Y HERRAMIENTAS EN UN PROYECTO BIG DATA



TECNOLOGÍA BIG DATA PARA LA PROSPECCIÓN Y DETECCIÓN DE NECESIDADES FORMATIVAS EN CLM

FUENTE DE DATOS



Fundación Estatal
PARA LA FORMACIÓN EN EL EMPLEO

FORMATO DE DATOS



ALMACENAMIENTO



CONSULTORÍA SEMÁNTICA

Ocupaciones

Habilidades/Competencias

Cualificaciones

TRANSFORMACIÓN



MODELAJE

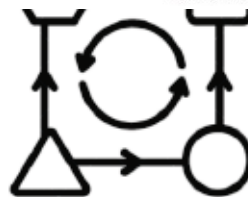
MODELAJE



CUADRO DE MANDO

Adaptación de la información rescatada a los objetivos previstos por la entidad que requiera los datos, ejemplos :

- Creación de formaciones para puestos de trabajo emergentes
- Políticas de empleo
- Actualización de competencias
- Ajuste por territorio y sector



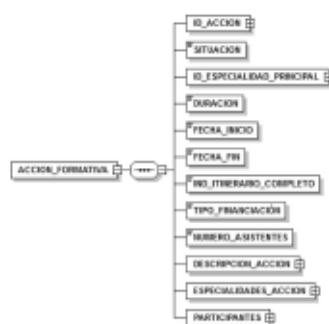
VISUALIZACIÓN

Representar los datos de manera fácilmente comprensible



ACCIONES FORMATIVAS

Representar los datos de manera fácilmente comprensible



Fuente: Elaboración propia

2. Infografía nº2: “Uso de la tecnología Big Data a nivel nacional y regional”



Fuente: Elaboración propia

II. INFORME EJECUTIVO

INFORME EJECUTIVO

PE/2020/10 - "PROSPECCIÓN Y DETECCIÓN DE NECESIDADES FORMATIVAS A TRAVÉS DE LA TECNOLOGÍA BIG DATA EN CLM"



RESULTADOS

INTRODUCCIÓN

El presente proyecto surge de la necesidad de **desarrollo y operativización de un sistema de prospección de necesidades formativas a través del uso de las tecnologías de la información, en este caso de la tecnología Big Data**. la importancia de desarrollar un sistema eficaz de detección y prospección de necesidades formativas, logrando que estas se ajusten a los requerimientos del sistema productivo y se anticipen a los cambios del mismo, y de esta forma, contribuir al desarrollo profesional y personal de los trabajadores, así como a la mejora competitiva de las empresas. El uso de un sistema de análisis de datos a gran escala es clave para alcanzar este objetivo, y mejorar las estrategias de formación para el empleo.

CONCLUSIONES

Para cumplir con los objetivos del proyecto es necesario que se cuente con unos KPI's claros, derivados de una consultoría previa, ya sea interna como externa. Esta es la parte más importante.

KPI's planteados (preguntas a responder por el proyecto):

- Competencias laborales más demandadas en el mercado laboral
- Ocupaciones con más (y menos) ofertas de empleo
- Ofertas según espacio de tiempo
- Ofertas según zona geográfica
- Tendencias
- Cursos más ofertados
- Relación entre formación y ofertas de empleo, correspondencia

Equipo de trabajo:

- Dirección técnica del proyecto con experiencia en gestión y coordinación de proyectos con tecnología Big Data, y conocimiento en dirección
- Analistas (data science) con experiencia en proyectos de Big Data, sería recomendable que tenga conocimientos sobre formación
- Arquitecto de datos que construya la herramienta y su interfaz
- Consultor de estrategia de proyecto, que plantee y marque los KPI's a desarrollar en la puesta en marcha de la herramienta

Las fuentes de datos sugeridas son:

Para ofertas de empleo

- Sepecam
- Indeed
- Infojobs
- LinkedIn
- Beebe

Para obtener datos sobre formación para el Empleo:

- Fundae
- Sepe
- Consejería de Empleo de Castilla-La Mancha (oferta de formaciones de todas las modalidades)

Herramientas:

- Apache Handcoop
- SQL
- R
- Python
- Tableau
- Cloudera

Protección de datos.

Acuerdos de seguridad por la anonimización de datos, y acuerdos marcos con las entidades que se vayan a extraer datos para su explotación. Al tratar los datos para obtener información las páginas web deben ser consultadas.

Fuente: Elaboración propia

III. NEWSLETTER

PE_2020_10 PROSPECCIÓN Y DETECCIÓN DE NECESIDADES FORMATIVAS A TRAVÉS DE LA TECNOLOGÍA BIG DATA EN CLM



Tecnología Big Data

El uso de este tipo de tecnologías se extiende más allá del sector privado. La administración pública integra desde hace años este tipo de sistemas como herramientas útiles para la adopción de decisiones y articulación de estrategias que se adapten y ajusten en mayor medida a la realidad. Como ejemplo de ello, el tratamiento de la crisis sanitaria del Covid-19.

En los últimos meses, se han procesado, tratado y cruzado miles de datos para estudiar la pandemia desde multitud de espacios, tanto públicos como privados, para conseguir así una panorámica de la situación logre anticiparse en la toma de decisiones para su detención. El uso de las aplicaciones relacionadas con la analítica de datos es inmenso, ofreciendo conocimiento de valor al conectar, procesar e interpretar datos de manera continua e inmediata.

FASES Y HERRAMIENTAS DE UN PROYECTO BIG DATA



Datos crudos	Almacenamiento	Transformación
Modelación	Visualización	Toma de decisiones

PROYECTOS TERRITORIALES DE BIG DATA EN ESPAÑA Y CLM

En España:

- **Estudio de movilidad con tecnología Big Data:** Este estudio tenía como objetivo identificar la movilidad a nivel nacional, de comunidad autónoma, provincial y local, como apoyo a las labores de seguimiento de la evolución de las enfermedades, para evaluar la efectividad de las medidas de restricción de la movilidad adoptadas.
- **Proyecto Dataestur:** Big Data en el turismo. Un proyecto que mediante una plataforma agrupa datos turísticos de diferentes fuentes, como el INE, Renfe o el Banco de España.

En Castilla-La Mancha:

Proyecto Sapiens: Big Data para la salud
El SESCAM ha puesto en marcha este programa basado en el manejo de la información a gran escala para leer historias clínicas a partir del lenguaje natural y transformarlas en datos para generar bases que ofrezcan información fiable sobre el estado de salud de la población de CLM.



IV. CRONOGRAMA DE DIFUSIÓN

En este apartado se presenta la planificación del cronograma de difusión a emplear como herramienta para controlar la correcta divulgación entre las personas y agentes interesados de los productos finales generados en el marco del estudio: las infografías, correspondiente proyectos Big Data en el territorio y explicación de sus fases a desarrollar y herramientas. Por último, se presenta el informe ejecutivo, que resume las principales conclusiones del estudio.

El objetivo de la difusión es, por un lado, fomentar y sensibilizar sobre el futuro del sector, uno de las áreas protagonistas de los próximos años, y sus necesidades formativas, existentes tanto en las actividades que le conforman, como en los puestos de trabajo que necesita de la capacitación en estas herramientas tecnológicas para su actualización.

La divulgación de los productos finales se realiza a través de medios telemáticos como el correo electrónico y redes sociales, y es trasladada y difundida ante la Secretaría de Formación para el Empleo para ponerlos a disposición de CCOO CLM.

FECHA DE DIFUSIÓN	PRODUCTO FINAL	TARGET	OBJETIVO DE LA DIFUSIÓN
30 de septiembre de 2021	Infografías	Secretaría de Formación para el Empleo de CCOO CLM	Trasladar la información más relevante y significativa del estudio sobre las fases de un proyecto Big Data con foco en la detección de necesidades formativas, así como ejemplos de uso de esta tecnología en el territorio nacional y regional.
30 de septiembre de 2021	Informe ejecutivo y newsletter	Secretaría de Formación para el Empleo de CCOO CLM	Ofrecer la información de interés y conclusiones del estudio sobre las conclusiones extraídas para el desarrollo de una tecnología Big Data con los KPI's de detección de necesidades formativas en CLM.
Permanente	Infografía	Población general de la región, Administración, empresas y centros de formación	Trasladar la información más relevante y significativa del estudio sobre las fases de un proyecto Big Data con foco en la detección de necesidades formativas, así como ejemplos de uso de esta tecnología en el territorio nacional y regional.
Permanente	Informe ejecutivo y newsletter	Población general de la región, Administración,	Ofrecer la información de interés y conclusiones del estudio sobre las

		empresas y centros de formación	conclusiones extraídas para el desarrollo de una tecnología Big Data con los KPI's de detección de necesidades formativas en CLM.
--	--	---------------------------------	---

V. ACTIVIDADES REALIZADAS

OBJETIVOS PREVISTOS EN EL DESARROLLO DE LA FASE	GRADO DE CUMPLIMIENTO
Diseñar las infografías en formato digital dedicadas a la presentación del uso de la tecnología Big Data y su contextualización a nivel nacional y regional	Completo
Elaborar productos que definan de manera precisa los pasos a seguir para el correcto desarrollo de un proyecto de analítica de datos a través de la tecnología Big Data	Completo
Redactar informe ejecutivo con las conclusiones principales alcanzadas durante el estudio	Completo

ACTIVIDADES REALIZADAS	FECHA INICIO	FECHA FIN
Creación y edición de infografías y/o mapas conceptuales que representen la situación y tendencias del sector de las energías limpias en la región de Castilla-La Mancha	02/08/2021	30/09/2021
Diseño de itinerarios sintéticos que muestren las necesidades formativas y de recualificación de los trabajadores/as del sector	02/08/2021	30/09/2021
Elaboración de un informe ejecutivo con los principales resultados del estudio	02/08/2021	30/09/2021
Difusión a través de las vías de comunicación que se estipulen convenientes	02/08/2021	30/09/2021

BIBLIOGRAFÍA DE TODAS LAS FASES

7 herramientas Big Data para tu empresa. Recuperado de <https://www.iic.uam.es/innovacion/herramientas-big-data-para-empresa/>

Analizando el estado presente y la evaluación de los datos. (s.f.) Recuperado de <https://www.hiberus.com/tecnologias-diferenciales/data-analytics>

Anchiraico Trujillo, J.C. (2017) *Diseño de una arquitectura Big Data para la predicción de crisis en el trastorno bipolar*. UCM. Madrid. Recuperado de <https://eprints.ucm.es/id/eprint/41633/1/tfm-bigdatapcbip%20-%20julioanchiraico.pdf>

Barea Mendoza, J. et al. (2018) *Oportunidades y retos de los macrodatos en la toma de decisiones sanitarias*. Recuperado de <https://fundaciongasparcasal.org/wp-content/uploads/2021/01/oportunidades-y-retos-de-los-macrodatos.pdf>

Barry, C. (1999). *Las habilidades de información en un mundo electrónico: la formación investigadora de los estudiantes de doctorado*. Anales de Documentación. Universidad de Murcia. Recuperado de <https://revistas.um.es/analesdoc/article/view/2731/2691>

Becerra, G. (2018) *Big data como objeto de estudio y método para la investigación empírica en sociología y psicología social*. STS, Simposio Argentino sobre Tecnología y Sociedad

Big Data: desde los inicios hasta hoy (s.f.) "Vodafone: Ideas para tu empresa". Recuperado de <https://ideasparatuempresa.vodafone.es/big-data-desde-los-inicios-hoy/>

Cedefop (2019) *Online job vances and skills analysis*. Recuperado de <https://www.cedefop.europa.eu/en/publications-and-resources/publications/4172>

Cómo es el almacenamiento de datos en tiempos del Big Data (2019) Recuperado de <https://blogs.imf-formacion.com/blog/tecnologia/como-es-el-almacenamiento-de-datos-en-tiempos-del-big-data-201907/>

Data Lake vs Data Warehouse. (5 de abril de 2021) Recuperado de <https://luminousmen.com/post/data-lake-vs-data-warehouse>

Descubre por qué los Big Data y Hadoop parecen ser inseparables. (s.f.) Tableau. Recuperado de <https://www.tableau.com/es-mx/learn/articles/big-data-hadoop-explained>

Diccionario económico El Economista. Recuperado de <https://www.eleconomista.es/diccionario-de-economia/big-data#:~:text=Denominamos%20Big%20Data%20a%20la,gesti%C3%B3n%20y%20procesamiento%20de%20datos>

Estudio de movilidad con Big Data (s.f.) Recuperado de <https://www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data>

Estudio: PE/2019/02 “Asesoramiento a pymes sobre la aplicación de la tecnología Big Data en la formación de sus trabajadores y la creación de empleo”. Elaborado por FORECLM y CCOOCLM y financiado por la Junta de Comunidades de Castilla-La Mancha.

Forem CLM-CCOO CLM (2020) *Asesoramiento a PYMES sobre la aplicación de la tecnología Big Data en la formación de sus trabajadores y la creación de emple*”.

Fundae “*Formación en las empresas - Castilla-La Mancha 2019*”, Recuperado de <https://www.fundae.es/docs/default-source/publicaciones-y-evaluaciones/estudios/2019/formaci%C3%B3n-en-las-empresas-2019---castilla-la-mancha.pdf>

Fundéu BBVA (2019) Consulta “*Big Data*”

García, R. (11 de marzo de 2015) *Business Intelligence aplicado a las Políticas Activas de Empleo en el área de Formación*. Recuperado de <https://www.hiberus.com/crecemos-contigo/business-intelligence-aplicado-las-politicas-activas-de-empleo-en-el-area-de-formacion/>

Instituto Nacional De Estadística (2020) *Encuesta de uso de TIC y Comercio Electrónico*.

La importancia de la visualización de los datos en Big Data (25 de agosto de 2020) Recuperado de <https://opensistemas.com/la-importancia-de-la-visualizacion-de-datos-en-big-data/>

Ley 30/2015, de 9 de septiembre, por la que se regula el *Sistema de Formación Profesional para el empleo en el ámbito laboral*. 217, de 10 de septiembre de 2015, páginas 79779 a 79823 (45 págs.) Recuperado de https://www.boe.es/diario_boe/txt.php?id=BOE-A-2015-9734

Los científicos alertan: España será un país de analfabetos si no invierte en I+D (18 de marzo de 2021) Recuperado de <https://murciaeconomia.com/art/77900/los-cientificos-alertan-espana-sera-un-pais-subalterno-si-no-invierte-en-id>

Los datos del Big Data. (s.f.). Campus Big Data. Entrada Blog. Recuperado de <https://www.campusbigdata.com/big-data-blog/item/111-datos-big-data>

Real Academia Española. (2019). *Diccionario de la lengua española*

Mayer-Schönberger,V y Cukier,K (2013) Big data: La revolución de los datos masivos. Colección Noema. Recuperado de <http://catedradatos.com.ar/media/3.-Big-data.-La-revolucion-de-los-datos-masivos-Noema-Spanish-Edition-Viktor-Mayer-Schonberger-Kenneth-Cukier.pdf>

Mota, C (2014) “Big Data. Un nuevo paradigma de análisis de datos” Recuperado de <https://www.iit.comillas.edu/docs/IIT-14-153A.pdf>

Morcillo, F (10 de julio de 2019) “KLa UCLM abre un centro de Big Data y Ciberseguridad en Albacete” ABC. Recuperado de https://www.abc.es/espana/castilla-la-mancha/abci-uclm-abre-centro-data-y-ciberseguridad-albacete-201907101436_noticia.html

Qué es el Big Data (s.f.). Recuperado de <https://www.auraquantic.com/es/que-es-big-data/>

¿Qué es un data lake? Definición, creación y ejemplos (18 de septiembre de 2019) Recuperado de <https://www.itainnova.es/blog/big-data-y-sistemas-cognitivos/que-es-data-lake-definicion-creacion-y-ejemplos/>

¿Qué es la transformación de datos? (s.f.) Tecnologías De La Información Recuperado de <https://www.tecnologias-informacion.com/transformacion.html#:~:text=La%20transformaci%C3%B3n%20de%20datos%20es,un%20nuevo%20sistema%20de%20destino.>

Seis consejos para optimizar la visualización de datos en la empresa (s.f.) Recuperado de <https://haycanal.com/noticias/6868/seis-consejos-para-optimizar-la-visualizacion-de-datos-en-las-empresa>

Statista (2020). Estadísticas Sobre El Uso Del Big Data.

VALLES (1997). Técnicas cualitativas de investigación social. Reflexión metodológica y práctica profesional (Madrid: Síntesis) Capítulo 6: “Técnicas de conversación, narración (I): Las entrevistas en profundidad”

TRABAJO DE CAMPO: HERRAMIENTAS DE ANÁLISIS E INDICADORES DE VALORACIÓN METODOLÓGICA

Se aplica una metodología cualitativa mediante dos técnicas diferentes: la entrevista semiestructurada y la reunión de trabajo con expertos en tecnologías Big Data y de ejecución de proyectos en la rama de las nuevas tecnologías (datos en diferentes sectores productivos). Estos profesionales, dada la naturaleza del método de investigación, proporcionan a través de su *expertise* información relevante para cubrir los objetivos de estudio, concretamente sobre las fases de un proyecto big data, las fuentes de datos adecuadas, así como las herramientas de calidad en este proceso. En líneas generales, se abarcó toda la información que fuera de para la propuesta de una línea de trabajo sobre estos puntos para poner en marcha una herramienta de esta tipología para la detección de necesidades formativas.

Se realizaron 3 entrevistas, 2 de ellas con personas expertas consultadas en el proyecto anterior (PE/2019702) y un grupo de trabajo con 3 personas con responsabilidad en la ejecución de proyectos.

En cada una de las fases desarrolladas, excepto en la Fase 4 de difusión donde se expone el conocimiento y la información ya generada en las fases previas, se emplean diferentes indicadores de valoración metodológica para garantizar la calidad de la investigación. En la Fase 2 se lleva a cabo el trabajo de campo a través de la reunión de expertos y las entrevistas y en la Fase 3 el análisis cualitativo de los datos recogidos a través de la metodología establecida.

Indicadores de valoración metodológica: Fase 2 y 3	Trabajo de campo: Fase 2	Herramientas de análisis: Fase 3
✓ La fase de investigación documental desarrollada permite conocer el contexto en el que se enmarca el estudio, y disponer de conocimientos sobre la materia para enfocar correctamente las cuestiones metodológicas en las sucesivas fases ✓ Gestión y selección adecuada de los perfiles a entrevistar necesarios para el cumplimiento de los objetivos previstos. ✓ Gestión y selección adecuada de los perfiles necesarios para el grupo de trabajo. ✓ Elaboración del guion de moderación para el adecuado desarrollo de del grupo ✓ Elaboración del guion de moderación para el adecuado desarrollo de las entrevistas cualitativas ✓ Análisis discurso de los participantes del grupo de trabajo y de las entrevistas a los distintos perfiles participantes en el trabajo de campo, asociando su tipo de conocimiento experto a los objetivos de estudio ✓ Grabación en audio de la conversación abordada	✓ Realización de reunión de trabajo con expertos/as ✓ Entrevista a perfiles expertos/as: ✓ Tecnología Big Data ✓ Conocimiento en formación ✓ Seguimiento y control del trabajo de campo	✓ Análisis de contenido del grupo de trabajo y las entrevistas en materia tecnología Big Data con objeto de la detección de necesidades formativas en la región de Castilla-La Mancha. El análisis de contenido se apoya en <i>verbatim</i>s ✓ Estudio cualitativo de las narraciones y lenguaje empleado por los informantes clave a través del análisis del discurso.

RELACIÓN DE PARTICIPANTES

Perfil	Nombre y Apellidos
Directora General	Encarnación Sánchez Vega
✓ Diseño del proyecto ✓ Seguimiento del desarrollo del proyecto ✓ Dirección del proyecto ✓ Selección y gestión de recursos humanos ✓ Gestión económica del proyecto ✓ Control y evaluación de la calidad en las diferentes fases de los proyectos ✓ Realización del análisis de la viabilidad del proyecto ✓ Toma de decisiones en relación con la evaluación interna del proyecto ✓ Gestión del riesgo ✓ Motivación al equipo ✓ Gestión de los conflictos que se puedan presentar	

Perfil	Nombre y Apellidos
Coordinadora técnica del proyecto	María Blanca de Pablos Rujas
✓ Coordinación técnica del proyecto y del equipo asociado al proyecto ✓ Confección, actualización y monitorización del plan de proyecto. ✓ Desarrollo el plan del proyecto ✓ Seguimiento y gestión global del proyecto y de sus fases. ✓ Planificación de recursos técnicos, humanos, materiales y económicos del proyecto. ✓ Seguimiento y control presupuestario ✓ Validación técnica de las actividades ejecutadas por los técnicos del proyecto ✓ Diseño y seguimiento de la certificación y justificación económica del proyecto ✓ Ejecución de acciones correctoras cuando sea necesario. ✓ Interlocución con la Junta de Comunidades de Castilla- La Mancha, y equipo de trabajo. ✓ Gestión de las compras y los proveedores. ✓ Elaborar los informes económicos de los proyectos, asegurando el cumplimiento de la normativa y procedimientos	

Perfil	Nombre y Apellidos
Técnico	Esther Miota
✓ Revisión bibliográfica y explotación de datos secundarios. ✓ Diseño metodológico investigación cualitativa a través de entrevistas y grupos de discusión. ✓ Trabajo de campo, entrevistador y moderador ✓ Aplicación del diseño metodológico de investigación cuantitativa a través de las técnicas empleadas en el proyecto (muestreo, operacionalización, cuestionarios, pre-test...) ✓ Uso de programas especializados para la encuestación y tratamiento de la información cualitativa ✓ Análisis cualitativo ✓ Análisis cuantitativo ✓ Planificación de informes parciales y finales ✓ Realización del trabajo de campo ✓ Moderación de los grupos de discusión ✓ Elaboración de informes con las conclusiones y recomendaciones del estudio	

Perfil	Nombre y Apellidos
Técnico	Amparo Herreros
✓ Revisión bibliográfica y explotación de datos secundarios. ✓ Diseño metodológico investigación cualitativa a través de entrevistas y grupos de discusión. ✓ Trabajo de campo, entrevistador y moderador ✓ Aplicación del diseño metodológico de investigación cuantitativa a través de las técnicas empleadas en el proyecto (muestreo, operacionalización, cuestionarios, pre-test...) ✓ Uso de programas especializados para la encuestación y tratamiento de la información cualitativa ✓ Análisis cualitativo ✓ Análisis cuantitativo ✓ Planificación de informes parciales y finales ✓ Realización del trabajo de campo ✓ Moderación de los grupos de discusión ✓ Elaboración de informes con las conclusiones y recomendaciones del estudio	

Perfil	Nombre y Apellidos
Técnico	Oscar García
✓ Revisión bibliográfica y explotación de datos secundarios. ✓ Diseño metodológico investigación cualitativa a través de entrevistas y grupos de discusión. ✓ Trabajo de campo, entrevistador y moderador ✓ Aplicación del diseño metodológico de investigación cuantitativa a través de las técnicas empleadas en el proyecto (muestreo, operacionalización, cuestionarios, pre-test...) ✓ Uso de programas especializados para la encuestación y tratamiento de la información cualitativa ✓ Análisis cualitativo ✓ Análisis cuantitativo ✓ Planificación de informes parciales y finales ✓ Realización del trabajo de campo ✓ Moderación de los grupos de discusión ✓ Elaboración de informes con las conclusiones y recomendaciones del estudio	

Perfil	Nombre y Apellidos
Técnico	Jose María Martín Pose
✓ Revisión bibliográfica y explotación de datos secundarios. ✓ Diseño metodológico investigación cualitativa a través de entrevistas y grupos de discusión. ✓ Trabajo de campo, entrevistador y moderador ✓ Aplicación del diseño metodológico de investigación cuantitativa a través de las técnicas empleadas en el proyecto (muestreo, operacionalización, cuestionarios, pre-test...) ✓ Uso de programas especializados para la encuestación y tratamiento de la información cualitativa ✓ Análisis cualitativo ✓ Análisis cuantitativo ✓ Planificación de informes parciales y finales ✓ Realización del trabajo de campo ✓ Moderación de los grupos de discusión ✓ Elaboración de informes con las conclusiones y recomendaciones del estudio	

Perfil	Nombre y Apellidos
Técnico	Nuria Conde
✓ Revisión bibliográfica y explotación de datos secundarios. ✓ Diseño metodológico investigación cualitativa a través de entrevistas y grupos de discusión. ✓ Trabajo de campo, entrevistador y moderador ✓ Aplicación del diseño metodológico de investigación cuantitativa a través de las técnicas empleadas en el proyecto (muestreo, operacionalización, cuestionarios, pre-test...) ✓ Uso de programas especializados para la encuestación y tratamiento de la información cualitativa ✓ Análisis cualitativo ✓ Análisis cuantitativo ✓ Planificación de informes parciales y finales ✓ Realización del trabajo de campo ✓ Moderación de los grupos de discusión ✓ Elaboración de informes con las conclusiones y recomendaciones del estudio	

Perfil	Nombre y Apellidos
Técnico de apoyo	Carolina Vidal
✓ Apoyo en la revisión bibliográfica y explotación de datos secundarios. ✓ Búsquedas avanzadas sobre la normativa, reglamentos, asociados al objeto de estudio ✓ Soporte en las labores de contactación con los expertos ✓ Apoyo en el trabajo de campo (base de datos de informantes clave, elaboración del guion para los grupos de discusión o entrevistas, notas de análisis de los grupos, transcripciones, selección de verbatims, exportación de los datos obtenidos a través de la encuesta a herramientas de análisis estadístico...) ✓ Soporte en la elaboración de gráficos, mapas de posicionamiento y conceptuales, contenidos o tablas con los datos recogidos durante el trabajo de campo	

Perfil	Nombre y Apellidos
Técnico	Encarnación Tarancón
✓ Apoyo en la revisión bibliográfica y explotación de datos secundarios. ✓ Búsquedas avanzadas sobre la normativa, reglamentos, asociados al objeto de estudio ✓ Soporte en las labores de contactación con los expertos ✓ Apoyo en el trabajo de campo (base de datos de informantes clave, elaboración del guion para los grupos de discusión o entrevistas, notas de análisis de los grupos, transcripciones, selección de verbatim, exportación de los datos obtenidos a través de la encuesta a herramientas de análisis estadístico...) ✓ Soporte en la elaboración de gráficos, mapas de posicionamiento y conceptuales, contenidos o tablas con los datos recogidos durante el trabajo de campo	

Perfil	Nombre y Apellidos
Técnico de apoyo	Miren Nekane
✓ Apoyo en la revisión bibliográfica y explotación de datos secundarios. ✓ Búsquedas avanzadas sobre la normativa, reglamentos, asociados al objeto de estudio ✓ Soporte en las labores de contactación con los expertos ✓ Apoyo en el trabajo de campo (base de datos de informantes clave, elaboración del guion para los grupos de discusión o entrevistas, notas de análisis de los grupos, transcripciones, selección de verbatim, exportación de los datos obtenidos a través de la encuesta a herramientas de análisis estadístico...) ✓ Soporte en la elaboración de gráficos, mapas de posicionamiento y conceptuales, contenidos o tablas con los datos recogidos durante el trabajo de campo	

Perfil	Nombre y Apellidos
Informático	Francisco Javier Gutierrez
✓ Definición y creación de soluciones técnicas en la tecnología del proyecto. ✓ Entender y ajustarse a la arquitectura del proyecto. ✓ Identificación y construcción de pruebas de desarrollo que cubran el comportamiento requerido de los componentes técnicos. ✓ Análisis de las necesidades del proyecto y desarrollo de soluciones de software. ✓ Soporte informático en todas las fases del proyecto ✓ Apoyar en la elaboración de informe de datos ✓ Elaboración y gestión de contenidos en formato video, multimedia, gráficos ... ✓ Diseño gráfico de carteles, panfletos, trípticos, infografías, logotipos, guías de usuario... ✓ Edición, gestión y administración de contenidos digitales y multimedia. ✓ Creación y gestión de contenidos para RRSS. ✓ Campañas de difusión en redes sociales ✓ Elaboración de contenidos en formato de vídeo, gráficos, post, banners multimedia para difusión en redes sociales	

Perfil	Nombre y Apellidos
Técnico de justificación económica	José María Martín Pose
✓ Realizar el seguimiento y justificación económico-administrativa de ayudas ✓ Revisar la normativa aplicable y validar de solicitudes y documentación asociada. ✓ Apoyar y asesorar en el proceso de solicitud de la subvención. ✓ Revisar y validar la documentación justificativa. ✓ Coordinarse con las áreas para la gestión de la justificación y obtención de la documentación acreditativa del gasto. ✓ Asesorar e informar a los técnicos del proyecto sobre los procedimientos internos a seguir para la eficiente justificación de los proyectos. ✓ Apoyar y asesorar al técnico de gestión de proyectos sobre las normativas a cumplir para la ejecución y justificación correcta de las ayudas concedidas. ✓ Contabilidad del proyecto: recepción de documentos contables, introducción de apuntes y archivo.	

Perfil	Nombre y Apellidos
Técnico de justificación económica	Oscar García
✓ Realizar el seguimiento y justificación económico-administrativa de ayudas ✓ Revisar la normativa aplicable y validar de solicitudes y documentación asociada. ✓ Apoyar y asesorar en el proceso de solicitud de la subvención. ✓ Revisar y validar la documentación justificativa. ✓ Coordinarse con las áreas para la gestión de la justificación y obtención de la documentación acreditativa del gasto. ✓ Asesorar e informar a los técnicos del proyecto sobre los procedimientos internos a seguir para la eficiente justificación de los proyectos. ✓ Apoyar y asesorar al técnico de gestión de proyectos sobre las normativas a cumplir para la ejecución y justificación correcta de las ayudas concedidas. ✓ Contabilidad del proyecto: recepción de documentos contables, introducción de apuntes y archivo.	

Perfil	Nombre y Apellidos
Administrativo	Eugenia Blanco
✓ Realización de las gestiones administrativas del proyecto ✓ Tramitación de documentos o comunicaciones internas o externas en los circuitos de información del proyecto. ✓ Elaboración de documentos y comunicaciones a partir de órdenes recibidas, información obtenida y/o necesidades detectadas en las diferentes fases del proyecto. ✓ Clasificación, registro y archivo de las comunicaciones y documentos según las técnicas apropiadas y los parámetros establecidos por el equipo técnico del proyecto. ✓ Tramitación y gestión administrativa en la presentación de documentos para la administración en plazo y forma requeridos. ✓ Proceso y registro de la información mediante archivos informáticos y bases de datos ✓ Manejo de estadísticas. ✓ Maquetación de informes y memorias de las diferentes fases del proyecto. ✓ Manejo de las aplicaciones ofimáticas en la gestión de la información y la documentación ✓ Tratamiento de datos y textos en terminales informáticos en condiciones de seguridad, calidad y eficiencia.	

CONCLUSIONES Y PROPUESTAS DE MEJORA

Investigación documental:

La escucha activa del mercado de trabajo en la región de Castilla-La Mancha es imprescindible para la permanente actualización de las políticas activas de empleo. Asimismo, en el contexto actual, con la transformación digital en permanente crecimiento y el avance de la digitalización en todos los ámbitos, hace necesario el replanteamiento de las cualificaciones y los cursos que forman al conjunto de personas trabajadoras de la región.

A través de este proyecto, se propone la creación una hoja de ruta para la configuración de un sistema Big Data que logre recolectar datos de diferentes fuentes de información (portales públicos, y privados), para detectar patrones de comportamiento del mercado de trabajo y de las competencias ligadas a él, de forma "regular y dinámica" y de esta forma se ajusten las formaciones y competencias a los requisitos del mercado y sus trabajadores/as.

El análisis documental de esta primera fase concluye con diversas propuestas de herramientas y pasos a seguir. Será en la fase de campo, gracias al conocimiento y experiencia de los informantes clave que se seleccionen, cuando se recojan los detalles técnicos correspondientes a la operativización de un proceso de datos como el que se pretende. Una de las conclusiones a las que se ha llegado, tras la revisión es la importancia de la combinación de fuentes, de donde surge un plus de valor. Los portales de empleo tienen un gran potencial, pero no son los únicos que contienen información sobre las demandas del mercado laboral, también los servicios públicos de empleo, y las consejerías pueden ofrecer datos que sirvan para su cruce.

Cada vez existen más herramientas de procesamiento de datos (depuración, análisis, visualización, transformación...), así como más opciones de almacenamiento. Esto plantea un horizonte positivo para la creación de este tipo de sistemas para entidades de mayor y menor tamaño, así públicas como privadas. Como ejemplo, el Servicio de Salud de Castilla-La Mancha (SESCAM) es pionero en integrar el sistema Big Data en conjunto con la Inteligencia Artificial. Gracias a este proyecto, se pretende acelerar el diagnóstico y la detección temprana de problemas.

Trabajo de campo:

El proceso de captación, dadas las características del presente proyecto que deriva de uno anterior, ha sido sencilla para la ejecución de las entrevistas, contando con expertos que participaron en las mesas virtuales y que estaban en contacto con la temática, además de poseer un conocimiento experto de gran valor.

La captación de los informantes clave para la reunión de trabajo se definió con un claro precepto: contar con expertos que tuviesen amplio conocimiento en creación y ejecución de proyectos. Así dos de ellos, además, tienen una relación directa con el análisis de datos. La predisposición de estos a ayudar para el proyecto se ha de destacar, así como la información de valor que aportaron.

La plataforma virtual empleada fue eficiente, desplazando al ADOBECONNECT que se utilizó el año anterior. El equipo se decantó por esta para facilitar el acceso a los informantes que iban a participar en la reunión, para no entorpecer la comunicación con instalaciones de software en sus equipos.

La tecnología Big Data, así como cualquier proceso digital, cambia muy rápidamente. Los programas sugeridos el año anterior el proyecto PE/2019/02, tienen ya sucesores. Se necesita una actualización constante de conocimiento.

Quedan claros puntos clave a profundizar en el análisis: Equipo de recursos humanos, perspectiva legal, acuerdos marco con administración y empresas de empleo como Infojobs (se pueden extraer los datos pero es necesario un acuerdo para la extracción de información con el procesamiento de estos).

Análisis de los resultados y su difusión:

Para cumplir con los objetivos del proyecto es necesario que se cuente con unos KPI's claros, derivados de una consultoría previa, ya sea interna con un consultor propio, como externa.

KPI's planteados:

- Competencias laborales más demandadas en el mercado laboral
- Ocupaciones con más (y menos) ofertas de empleo
- Ofertas según espacio de tiempo
- Análisis según zona geográfica
- Tendencias
- Cursos más ofertados
- Relación entre formación y ofertas de empleo, correspondencia

Equipo de trabajo:

- Dirección técnica del proyecto con experiencia en gestión y coordinación de proyectos con tecnología big data, y conocimiento en dirección
- Analistas con experiencia en proyectos de esta tipología
- Arquitecto de datos que construya la herramienta y su interfaz
- Consultor de estrategia de proyecto, que plantee y marque los KPI's a desarrollar

en la puesta en marcha de la herramienta

Las fuentes de datos sugeridas son:

1. Ofertas de empleo:

- Sepecam
- Indeed

- Infojobs
- LinkedIn
- RRSS
- ...

2. Formación para el Empleo:

- Fundae
- Sepe
- Consejería de Empleo de Castilla-La Mancha (oferta de formaciones de todas las modalidades)
- Entidades de formación

Otras consideraciones del proyecto son: costes y protección de datos.

Se han elaborado 2 infografías con las conclusiones del estudio para su difusión:

La primera infografía denominada “Fases y herramientas de un proyecto Big Data” donde se exponen tanto los pasos a seguir en la explotación y análisis de datos, como las herramientas clave a utilizar.

La segunda infografía denominada “Uso de la tecnología Big Data a nivel nacional y regional”, presenta información sobre el uso que presenta esta tecnología en el territorio a través de ejemplos

Por otro lado, se elabora un informe ejecutivo diseñado en formato digital con las principales conclusiones y retos del estudio.

La planificación del cronograma de difusión se presenta como herramienta para controlar la correcta divulgación de los productos finales generados en el marco del estudio: las infografías con los resultados del estudio y el informe ejecutivo. Esta planificación incluye el cronograma de difusión, el producto final a difundir, el target al que va dirigida la difusión y el objetivo de la difusión. En la planificación aparecen las vías disponibles de difusión que se prolongarán en el tiempo y se utilizarán cuando así lo requieran las circunstancias.

Siguiendo el planteamiento establecido en el Cronograma de difusión:

- Los productos finales fueron trasladados y difundidos ante la Secretaría de Formación para el Empleo para ponerlos a disposición de CCOO CLM.
- De manera permanente, y a través de las vías de divulgación que se decidan, se ponen a disposición los productos finales a la población de la región, la Administración, las empresas y los centros de formación cuando así lo requieran.

Se pone a disposición de la Dirección General de Formación Profesional para el Empleo de la Junta de Comunidades de Castilla-La Mancha de los productos finales para que, una vez revisado el estudio, se proceda a la divulgación del mismo tal como consideren

Propuestas de mejora:

Crear un grupo de trabajo con agentes sociales, administración y técnicos expertos en las tecnologías Big Data para la puesta en común de KPI's que permitan la creación de este proyecto de anticipación de las necesidades formativas del sistema productivo y sus trabajadores con los puntos clave extraídos en el presente proyecto. Contar con personal experto en la consultoría de proyectos de estas características.

Impulsar el **conocimiento digital de la analítica de datos** entre los agentes sociales y la administración para añadir valor a sus decisiones como parte de un posible grupo de trabajo. De esta forma estas se adecuarán mejor a la naturaleza del proyecto.

Apostar por una detección de necesidades para favorecer la empleabilidad a través de la adaptación de las necesidades formativas : la analítica de datos a través de la tecnología Big Data, tal como se vio en el proyecto PE.2019.02, es una herramienta que, si se plantea de forma correcta permitirá anticiparse a las necesidades de las personas trabajadoras y el sistema productivo de modo que sea más sencillo adaptar los currículos formativos o las políticas de empleo, según la situación de la región. Es un conjunto de arquitecturas tecnológicas que permite almacenar y manejar datos en grandes cantidades, para abastecer necesidades de negocio y ofrecer soluciones a problemas detectados. Su importancia va asociada a la puesta en valor que hace de los datos tras su explotación y análisis. De nada sirve la cantidad de datos ingentes que se pueden procesar, sino existe un correcto tratamiento a posteriori.

Dependiendo de la inmediatez del dato la configuración de la herramienta puede variar. Quizá, en un primer momento no sea necesaria conocer al momento como van cambiando contratos, formaciones y demás datos, simplemente con tener una foto mensual, o semanal sería suficiente, abaratando y simplificando la herramienta. Esta es una cuestión importante trasladada por las personas expertas en el momento de la creación de la herramienta.

Es importante que se haga una consultoría con personas expertas u organización dedicada ello, que ofrezca un buen asesoramiento y que analicen de forma adecuada los KPI's aquí planteados para establecer un trabajo analítico conciso que no genere costes adicionales.