

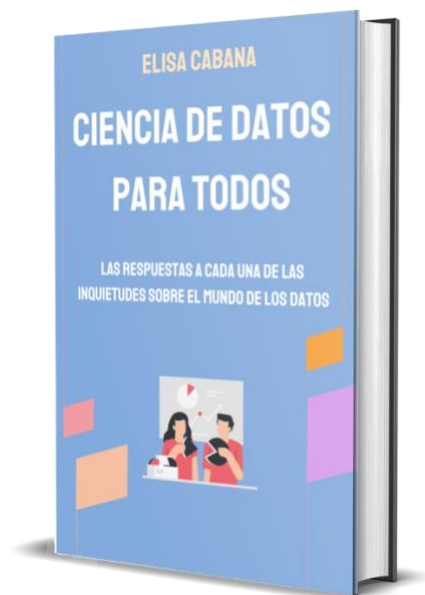
¿CÓMO CONVERTIRTE EN CIENTÍFICO DE DATOS?

Antes de comenzar, este libro es un pequeño resumen de algunas cuestiones importantes para aquellos que quieren comenzar una carrera en Ciencia de Datos. Son solamente 13 páginas pero te servirán como guía rápida para que tengas una idea de cómo es el proceso.

Ahora debo comentarte que, motivada por estas cuestiones y por las inquietudes de mis estudiantes, recientemente he lanzado un libro mucho más completo que abarca muchos más temas e inquietudes en detalle, incluyendo los temas más actuales relacionados con los datos: Big Data, Inteligencia Artificial, redes neuronales y sus aplicaciones, Chat-GPT, Copilot, DALL-E, Stable-Difussion, Computación cuántica, Visión por computador, y miles de tecnologías que están de moda hoy en día. Además incluye entrevistas a expertos científicos de datos que ya trabajan en este campo y compartirán su experiencia, su día a día y sus recomendaciones.

Aquí te dejo un enlace donde podrás ver más detalles sobre el libro y adquirirlo por un precio especial para estudiantes.

<https://aprendeconeli.com/libro>



Al final de este resumen encontrarás también el índice de los contenidos que incluye el libro para que te hagas una idea de lo que aprenderás.

LO QUE NECESITAS APRENDER

En esta sección vamos a ver cuáles son los pilares fundamentales de la Ciencia de Datos y cuáles son las habilidades complementarias que te van a ayudar en este camino.

LOS 3 PILARES DE LA CIENCIA DE DATOS

Para iniciar en el mundo de la Ciencia de Datos, existen 3 pilares fundamentales que constituyen el soporte del éxito en este campo, como se ilustra en el siguiente diagrama:



Los 3 pilares son: Estadística, Programación y Conocimiento de campo.

Estadística: Es la parte científica de lo que serán tus proyectos, lo que te enseñará la base y los métodos para saber cómo extraer información de los datos. Todos los que se quieren iniciar o están al inicio del proceso, tienen miedo a la parte de Estadística, sobre todo si vienen de un campo muy diferente, o han tenido una mala experiencia en el nivel escolar en las asignaturas de Matemática y Estadística. Sin embargo, con el tiempo verás que es un área del conocimiento muy bonita, que está presente en casi todos los aspectos del mundo real, que es muy necesaria y tiene muchísimas aplicaciones. Y la complejidad no es tanta como uno siempre se imagina. En Ciencia de Datos, no nos interesa tanto la estadística teórica en el sentido de crear un nuevo método sino que nos interesa la estadística aplicada, los métodos ya existentes que podemos aplicar en dependencia del problema que estemos resolviendo. Puede ser que tenga cierta complejidad el hecho de entender los métodos y cómo funcionan, pero esto es imprescindible para poder decidir luego en la práctica cuál debemos usar, cómo lo podemos configurar y cómo podemos interpretar los resultados. Pero es muy emocionante cuando ya tienes ese conocimiento y luego solo es cuestión de ponerlo en práctica para diferentes aplicaciones.

Programación: Tener conocimientos de programación es inevitable, porque en la práctica la mayoría de las veces no haremos los cálculos y modelos a mano, sino que haremos uso de herramientas que nos ayudan a que esos cálculos (que a veces son muy

complejos) resulten una tarea sencilla. Para Ciencia de Datos, en estos tiempos los dos lenguajes de programación más utilizados son R y Python. Hay otros complementarios, como SQL, Excel, etc., que quizás las empresas puedan pedir como requisito, pero que son muy sencillos y a veces lo puedes empezar a aprender cuando ya estás incorporado a la empresa. Sin embargo, los dos que mencioné anteriormente, Python y R son lenguajes un poco más avanzados, y son los que deberás aprender si quieres hacer análisis descriptivo de tus datos, análisis predictivo para predecir datos futuros en base a tus datos recogidos, aplicar métodos de machine learning (aprendizaje automático) o de deep learning (aprendizaje profundo).

Conocimiento de campo: Es la parte añadida y opcional sobre todo en los inicios del camino de la Ciencia de Datos, pero que a la larga también tiene una importancia significativa. Un detalle importante es que cada campo es diferente y tiene sus características, nociones y conceptos propios. Por ejemplo, si trabajas como analista de datos en un Banco, puede ser que seas muy bueno en Estadística y Programación, pero si no entiendes los conceptos detrás o relacionados con tu proyecto (por ejemplo, qué son las tasas de interés, o cómo funcionan las amortizaciones, etc.) será más complicado entregar un resultado significativo, válido y útil. Obviamente mientras más conocimiento tengamos del campo en el que estamos trabajando (finanzas, biotecnología, ingeniería, medicina, etc.), mejor entenderemos las preguntas y respuestas que podemos obtener de los datos y más fácil será la interpretación de los resultados. Sin embargo, al principio esto no es un requerimiento, sobre todo por el hecho de que no nos enfocaremos en un área de aplicación en concreto. Pero luego cuando ya te hayas incorporado a un campo específico, este será un pilar cada vez más importante, que irá reforzándose con el tiempo, con la misma práctica y experiencia.

OTRAS HABILIDADES COMPLEMENTARIAS

Para ser un buen científico o analista de datos, hay otro conjunto de habilidades que complementan a los pilares, pero que no son de menor importancia y que pueden ser claves a la hora de encontrar trabajo o de avanzar en tu propio trabajo. Estas habilidades son:

1. Habilidades de comunicación
2. Gestión de proyectos
3. Habilidades personales

Habilidades de comunicación: Si quieres que tus resultados y descubrimientos hagan una diferencia y sean reconocidos, tienes que saber comunicarlos y presentarlos a tu equipo, jefes o compañía. No tienes que ser el mejor comunicador oral del mundo, pero debes aprender a comunicar de manera segura y precisa cuando quieres explicar tus resultados.

Gestión de proyectos: La mayoría de los científicos de datos se benefician de cierta autonomía, sobre todo porque el resto de compañeros no entienden exactamente lo que hacemos. Sin embargo, esto no significa que puedes tardar un tiempo indefinido en entregar los resultados de tu problema o proyecto. Así que aquí lo importante es gestionar de manera eficiente tu propia productividad, cumplir con los plazos de entrega de resultados (deadlines), evaluar la importancia del proyecto, etc.

Habilidades personales: Los que trabajamos en este mundo de la ciencia de datos, la gran mayoría tendemos a ser introvertidos, pero a lo largo del tiempo, tenemos que aprender a

comunicarnos con las personas. Es importante entender que formaremos parte de un equipo, y si por ejemplo algo falla en los datos, tenemos que comunicarnos con los ingenieros de datos, o si descubrimos algo importante tenemos que comunicarlo a los compañeros expertos en el negocio o el campo específico en el que estamos. La comunicación debe ser siempre bilateral, porque los demás pueden aprender de ti y de tus resultados, pero tú también puedes aprender de los demás.

CURRÍCULUM

Una vez que has iniciado el camino de la Ciencia de Datos, y ya tienes los conocimientos básicos para empezar a buscar trabajo, hay una pregunta que siempre prevalece. ¿Cómo encontrar mi primer trabajo de científico de datos? No hay una respuesta única y efectiva para esta pregunta, pero sí hay algunos consejos que puedes seguir para aumentar tus posibilidades, y son los que te daré a continuación.

Lo primero que debes hacer es crear tu currículum, ya que esta es la principal carta de presentación a la hora de buscar trabajo. Es importante que este documento no solo refleje tu historia, sino también tus habilidades. Es decir, obviamente pondrás un resumen de quién eres, qué sabes y dónde lo has aprendido. Pero es muy importante remarcar que tienes la mayoría de los conocimientos que piden para ese empleo, y demostrarlo. Así como también remarcar y demostrar que tienes las habilidades complementarias que mencionamos antes.

Otro aspecto importante a la hora de buscar trabajo es que no siempre tienes que tener todos los conocimientos, pero los que son más importantes sí, que por supuesto son los que hemos mencionado antes, los pilares, sobre todo el de conocimientos de Estadística y de Programación.

A la hora de demostrarlo, está bien si haces referencia a tus estudios, ya sean carreras universitarias o cursos que has tomado por tu cuenta. En muchas ocasiones no es necesario en absoluto tener un título universitario, ni un máster, ni mucho menos un doctorado. El aprendizaje online está muy bien reconocido, y con tener un certificado que demuestre que has estudiado y profundizado en esos temas es suficiente. Lo que sí es importante es que tengas los conocimientos de verdad, porque en muchas ocasiones la empresa suele enviar tareas a resolver para filtrar a los solicitantes del empleo.

Sin embargo, el conocimiento no es lo único que deberías remarcar en tu currículum. Hay algunas cosas que puedes hacer, y que serán siempre positivas de cara a demostrar de otra manera tu conocimiento y tus habilidades. Una de ellas es crear repositorios abiertos (open repositories). La web más conocida de repositorios es GitHub: <https://github.com/>. En GitHub puedes crear un repositorio, o guardar/copiar repositorios de los demás que te puedan interesar. Lo más importante es que vayas guardando algunos de los códigos de análisis de datos que hayas hecho durante tu fase de aprendizaje, y que crees que demuestran tus conocimientos. Es sencillo, solo hay que registrarse, crear el repositorio y subir los códigos.

Otro consejo es tener una web personal. No tiene que ser de pago, hay muchos sitios donde puedes hacerla gratis: en GitHub, en Wordpress, etc. En esta web puedes incluir información sobre ti, noticias o entradas de blog de temas de Ciencia de Datos que te

interesen, tus propios análisis de datos, etc. Crear una web es una tarea que no es muy compleja porque hoy en día está todo muy automatizado, pero que te puede llevar algunas horas. Por ello te voy a recomendar una alternativa que creo que además es muy efectiva, y es tener un blog en la plataforma de Medium. Esta web es muy útil también para aprender, ya que muchos analistas y científicos de datos suben contenido constantemente en forma de entradas de blog escritas, y ahí explican mediante texto e imágenes, sus propios análisis y proyectos, la mayoría de las veces incluyendo el código y los datos. Pero eso también lo puedes hacer tú de cara a mostrar no sólo tus conocimientos sobre Estadística y Programación, sino también tus habilidades comunicativas y tu dedicación. Si quieres echar un vistazo a mi perfil en Medium, aquí te lo dejo, para que veas un ejemplo:

<https://ecab-estadistica.medium.com/>

LA RUTA DE APRENDIZAJE

Ahora que ya conoces los pilares fundamentales de la Ciencia de Datos, y las habilidades que te ayudarán a conseguir un empleo de científico de datos, puedes ir a la Ruta de Aprendizaje que hemos diseñado especialmente para adquirir los conocimientos necesarios en Ciencia de Datos: <https://aprendeconeli.com/ruta-de-aprendizaje/>

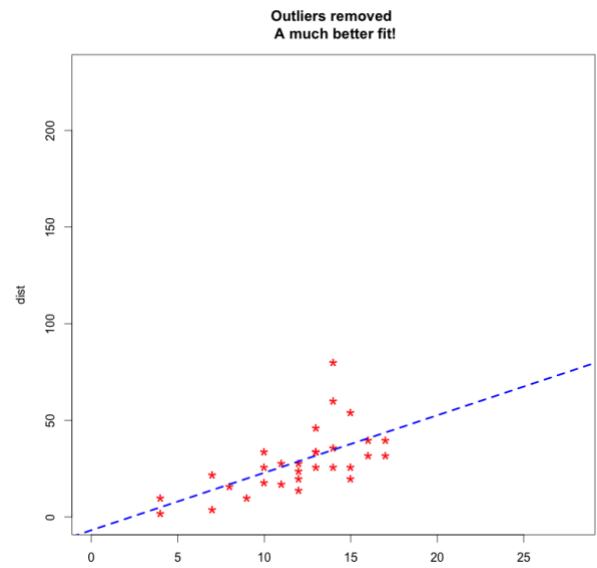
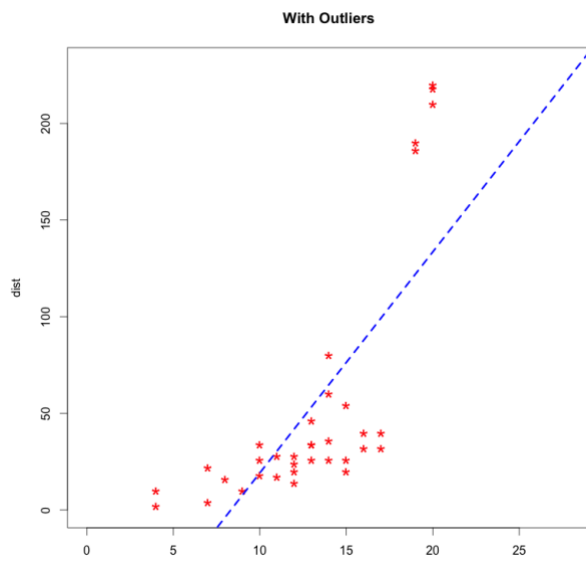
TIEMPO INVERTIDO

Lo primero que muchas veces mis estudiantes me preguntan, es: **¿cuánto tiempo debo invertir en este aprendizaje?**

La respuesta por supuesto depende de muchos factores. Por ejemplo, depende de la capacidad de aprendizaje de cada persona, hay quien le cuesta más y hay quien le cuesta menos. Depende también de los conocimientos previos que se tengan. Muchas veces ya se tiene una base de algunos de los conceptos iniciales porque se han tocado esos temas en los estudios anteriores de esa persona. Depende también del tiempo de dedicación, porque no es lo mismo dedicar todo el día, que 1-2 horas al día, o solo un par de horas a la semana. En promedio, puede tomar entre 3-6 meses si se dedican al menos 10 horas a la semana. Y aquí hay que remarcar que ese tiempo de aprendizaje debe ser tiempo de calidad, es decir, no debemos dejar que nos interrumpa ninguna distracción, y esto puede parecer trivial, pero es un aspecto clave.

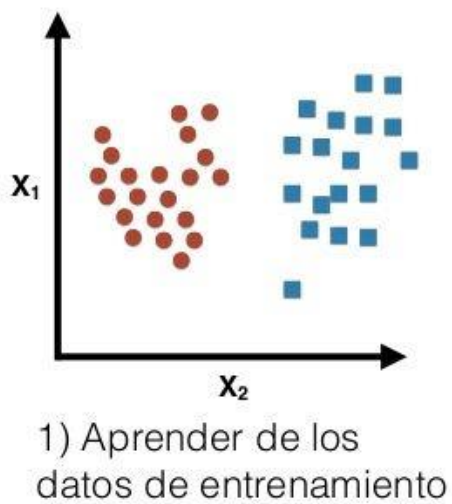
EJEMPLOS REALES

PREDICCIÓN DEL VALOR DE UNA VIVIENDA



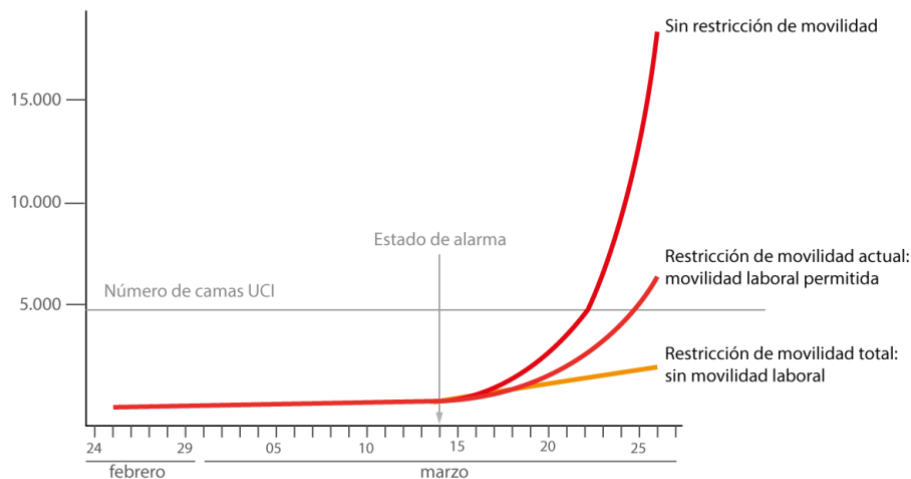
CLASIFICACIÓN PARA DETECTAR FRAUDE BANCARIO

Clasificación



PREDICCIÓN DE ENFERMOS DEL COVID

Predicción de casos críticos de coronavirus en España
Número de ingresos en UCI según el nivel las restricciones de movilidad



Fuente: Investigadores de la Universitat Rovira i Virgili y de la Universidad de Zaragoza

ABC

¿DÓNDE APRENDER?

UNIVERSIDAD VS. AUTOAPRENDIZAJE

Al contrario de lo que mucha gente piensa, no es necesario tener (ni mucho menos empezar) una carrera universitaria, ni un máster, ni un doctorado, para adquirir estos conocimientos. No obstante, las opciones más comunes y lo que cuesta en promedio en España son las siguientes:

1. **Carrera universitaria:** Suelen ser 4 años, con un gasto promedio de 2.000 euros por año.
2. **Master universitario:** Suelen ser de 1-2 años con un gasto promedio de 1.500 euros por año.
3. **Master privado en escuela de negocios:** Suelen ser de 1 año con un gasto promedio de 5.000-10.000 euros al año.

¿ES POSIBLE APRENDER DATA SCIENCE GRATIS?

Por supuesto. De hecho, si no necesitas un certificado, puedes hacer muchos de los cursos de EdX o Coursera de forma totalmente gratuita. No obstante, si quieres sacarles todo el partido y, además, poder acreditar tus conocimientos, merece la pena pagar la formación. Por otro lado, el problema de los cursos gratuitos de Data Science es que son principalmente de nivel básico o introductorio. Sin embargo, es una opción para ir introduciéndote en este mundo. Hay muchísimo material completamente gratuito, como por ejemplo: cursos gratis, vídeos en YouTube, blogs como Medium, webs como Kaggle, libros gratuitos, etc. Sin duda, el autoaprendizaje es hoy más accesible que nunca y está cada vez más al alcance de todos.

¿CUÁL ES LA GRAN VENTAJA DE LOS CURSOS DE CIENCIA DE DATOS YA ESTRUCTURADOS?

Una alternativa a estudiar en la universidad o hacer estudios en escuelas privadas o de negocios, es la educación online a través de cursos. Además de que los cursos suelen ser mucho más económicos que los programas de universidades o empresas de negocio, en los cursos estructurados tendrás toda la información bien organizada, graduada por nivel de dificultad, desde cero hasta el nivel más avanzado, y con un diseño mucho más atractivo. No tendrás que ir a buscar la información por tu cuenta sobre cada concepto, o cada método, sino que dispondrás de todo el contenido, tanto teórico como práctico, en un solo sitio. Plataformas como Udeemy son un buen punto de inicio porque además de ofrecer precios extremadamente económicos en comparación con el resto de propuestas, tienes un período de reembolso de 30 días por si no te ha gustado el contenido, tienes también la opción de preguntar dudas a los instructores del curso, tienes acceso ilimitado y de por vida para hacerlo a tu ritmo y cuando tengas tiempo, y puedes obtener un certificado válido de finalización cuando completes el curso.

RECOMENDACIONES

A continuación te dejo mis recomendaciones personales, en varios formatos para que tengas diferentes herramientas y puedas elegir la que mejor se adapte a ti.

CURSOS GRATUITOS:

Udeemy: [Curso gratis de ejercicios resueltos de Estadística](#) (gratis ilimitado)

Coursera: [Data Science, Statistics and Machine Learning](#) (gratis los primeros 7 días)

DATOS:

[Kaggle](#)

[Google Dataset Search](#)

CANALES DE YOUTUBE:

[Aprende con Eli](#)

[Dot CSV](#)

APLICACIONES PRÁCTICAS

[Medium](#)

[Towards Data Science](#)
[Kaggle](#)

CURSOS DE LA RUTA DE APRENDIZAJE DE APRENDE CON ELI

Esta es mi Ruta de Aprendizaje, son cursos que he ido creando basándome en los conocimientos más útiles, fundamentales y necesarios para emprender el camino de la Ciencia de Datos. Cada curso suele tener un precio entre 10-15 euros cuando están en descuento:

[Aquí puedes ver la ruta](#) y los cursos en el orden que sugiero:

1. Curso completo de Estadística a nivel universitario
2. Curso práctico de Estadística y Probabilidades
3. Curso práctico de Estadística Inferencial
4. Curso completo de Data Storytelling de cero a experto
5. Curso de R y Python para Data Science y Análisis de Datos
6. Curso avanzado de estadística no paramétrica con R y Python
7. Curso avanzado de estadística multivariante con R y Python
8. Curso de análisis avanzado de datos atípicos y outliers con R y Matlab
9. Curso avanzado de estadística bayesiana con Python
10. Curso avanzado de series temporales con R y Python
11. Curso de series temporales multivariantes con R y Python

PACKS DE CURSOS



[Por otro lado tienes disponibles 3 packs de cursos](#) en la Academia Aprende con Eli, para adquirir varios o todos los cursos, de una forma mucho más económica aún:

1. [Estadística de cero a experto](#): Tendrás acceso a 5 cursos individuales, más de 56 horas de contenido: estadística paramétrica, no paramétrica y bayesiana, más ejercicios resueltos. Todo en un solo curso completo.
2. [Data Science de cero a experto](#): 5 cursos individuales, más de 65 horas de contenido: R y Python, Multivariante, Atípicos, Series Temporales uni y multivariantes.

3. [Pack completo de Estadística y Data Science](#) (pack con todos los cursos juntos): 10 cursos individuales que incluyen los dos packs anteriores, más el curso gratis y los 2 Libros de Regalo. (+120h de contenido). Este pack tiene incluido un descuento adicional del 20% en el precio total.

LIBROS

¡Sumérgete en la Ciencia de Datos con mi nuevo libro “*Ciencia de datos para todos: Las respuestas a cada una de las inquietudes sobre el mundo de los datos*”! Descubre cómo comenzar una carrera en este emocionante campo, explora las últimas tendencias de la industria y recibe consejos directamente de los expertos que ya se dedican a esta disciplina. ¡Obtén tu copia ahora mismo! Disponible en versión digital y física.

<https://aprendeconeli.com/libro>



Tabla de Contenido del Libro

DEDICATORIA	v
AGRADECIMIENTOS	vii
Prólogo 1	
Por qué escribí este libro (y por qué deberías leerlo)	3
Mi historia	4
¿Por qué este libro es para ti?	10
Introducción a la ciencia de datos 13	
¿Qué es la ciencia de datos?	14
¿Qué tipos de datos existen según su estructura?	17
Historia y evolución de la ciencia de datos	19
¿Por qué es importante la ciencia de datos hoy en día?	21
Machine Learning, Inteligencia Artificial y Big Data	23
Aprendizaje Automático	24

Inteligencia Artificial	30
Big Data	33
Ingeniero, Analista y Científico de Datos	34
¿En qué se diferencian estas profesiones?	34
Veamos un proyecto de Ciencia de Datos	36
Salario y crecimiento de empleo	38
¿Qué tienen en común estas profesiones?	41
Fundamentos de la ciencia de datos	43
Estadística	43
Población y muestra	44
Variable aleatoria	45
Inferencia	46
Ejemplo de aplicación de la Estadística en salud	47
Programación	49
El lenguaje R	50
El lenguaje Python	52
Conocimiento de campo	56
Habilidades complementarias	57
Habilidades de comunicación	57
Habilidades de pensamiento crítico	59
Habilidades de aprendizaje autónomo	60
Habilidades de negocio	62
Habilidades de liderazgo	63
Proactividad	63
Creatividad	63
Habilidades de colaboración	64
Herramientas y técnicas de la ciencia de datos	65
Análisis descriptivo o exploratorio de datos	66
Tablas	66
Gráficos para visualizar los datos	69
Resúmenes estadísticos	75
Estadística paramétrica	77
Estadística no paramétrica	79
Datos multivariantes y aprendizaje automático	82
Aprendizaje supervisado	84
Aprendizaje no supervisado	89
Aprendizaje profundo y redes neuronales	91
Datos que causan problemas: los atípicos	92
El enfoque bayesiano	95
Los datos que dependen del tiempo: las series temporales	101
Fases de un proyecto de ciencia de datos	107

Casos de estudio y aplicaciones	115
Aplicaciones en redes sociales	116
Tipos de datos que se pueden obtener	116
Herramientas y algoritmos	117
Utilidad de este tipo de datos	118
Debate ético sobre los datos personales	119
Aplicaciones en medicina y salud	120
Innumerables beneficios	120
Detección de enfermedades	121
La ciencia de datos para el COVID-19	124
Ciencia de datos con fines sociales	128
Medicina personalizada	129
Aplicaciones en ciberseguridad y finanzas	130
La seguridad en el entorno cibernético	130
La bolsa y los mercados financieros	132
Finanzas personalizadas	133
Compañías de seguros	134
Análisis de riesgo	135
Detección de fraudes financieros	136
Aplicaciones en otras industrias y en la sociedad	137
Ganadería y Agricultura	137
Marketing	139
Maquinaria	140
Energía	142
Transporte y coches autónomos	142
Meteorología y cambio climático	144
Ética y privacidad en la ciencia de datos	147
Desafíos éticos y legales	148
Sesgo en los algoritmos de ML	148
Protección de datos personales y privacidad	153
Leyes de privacidad	153
Herramientas	154
Responsabilidad social en el uso de datos	156
El peligro de los datos: casos reales	157
Actualidad y Futuro de la ciencia de datos	161
Nuevas tendencias y tecnologías	162
Aprendizaje profundo avanzado	162
Internet de las cosas	168
Análisis de grafos	170
Computación en la nube	171
Automatización del aprendizaje automático	171
Privacidad diferencial	172

Inteligencia artificial explicativa	173
Las mejores herramientas de IA online	174
Impacto potencial en la sociedad	178
Oportunidades y desafíos para la investigación futura	181
Recomendaciones y consejos prácticos	183
Entrevista a expertos y sus recomendaciones	184
Experto en análisis de datos de movilidad para proyectos de investigación	184
Experto en visualizaciones de datos y creación de informes interactivos para empresas	189
Experto en Ciencia de Datos aplicada a la creación de mapas	192
Experta en Ciencia de Datos en el mundo bancario y financiero	194
Consejos prácticos para tener éxito en ciencia de datos	198
recursos adicionales	211
Conclusiones	217
ACERCA DE LA AUTORA	223