

Fundamentos de Ciencia de Datos con R

Gema Fernández-Avilés y José-María Montero

2023-05-18

Índice general

Prefacio	5
¡Hola mundo!	5
¿Por qué este libro?	6
¿A quién va dirigido?	7
El paquete CDR	8
¿Por qué R?	8
Agradecimientos	9
1. Ética en la ciencia de datos	11
1.1. ¿Qué es la ética?	11
1.2. Los principios éticos	12
1.3. Equidad: la importancia de los sesgos	14
1.4. ¿Es necesaria la explicabilidad?	16
1.5. Recursos en R para trabajar en sesgos y explicabilidad	18

Prefacio

¡Hola mundo!

El siglo XXI está siendo testigo de grandes cambios vertiginosos en el contexto social y tecnológico, entre otros. Los tiempos han cambiado, la sociedad se ha globalizado y “exige” respuestas inmediatas a problemas muy complejos. Vivimos en el mundo de la **información**, de los **datos**, o mejor, de las **bases de datos masivas**, y los ciudadanos y, sobre todo, las empresas y los gobiernos, dirigen su mirada hacia el mundo científico para que les ayude a “**oír las historias**” que cuentan esos datos acerca de la realidad de la que han sido extraídos. Y, dado su enorme volumen y sofisticación (en el nuevo mundo las imágenes y las frases, por ejemplo, también son datos) exigen algoritmos de nueva generación en el campo del *machine learning* o incluso del *deep learning* para “**oír las historias**” que cuentan. No parecen mirar al “antiguo” investigador científico, sino al “nuevo” *data scientist*.

Ello, inevitablemente, se traduce en la necesidad de profesionales con una gran capacidad de adaptación a este nuevo paradigma: los científicos de datos o “nuevos hombres del Renacimiento”, para lo cual las Universidades y demás instituciones educativas especializada se apresuran a incluir el grado de Ciencia de Datos en su oferta educativa y a ofertar seminarios de software estadístico de acceso abierto para sus estudiantes de primeros cursos.

Con la emergencia de la nueva sociedad, en la que el manejo de la ingente cantidad de información que genera se hace absolutamente necesario para circular por ella, la **Ciencia de Datos** ha venido para quedarse. Sin embargo, el mundo de la Ciencia de Datos es cualquier cosa menos sencillo. En él cualquier ayuda, cualquier guía es bienvenida. Por ello, es muy recomendable que la persona que se quiera introducir en él, con fines de investigación o con fines profesionales, se agarre de la mano de un guía especializado que le lleve, de una manera amena, comprensible y eficiente, desde el planteamiento de su problema y la captura de la información necesaria para poderle dar una solución, hasta la redacción de las conclusiones finales que ha obtenido con los modernos informes reproductibles colaborativos. Y como en la parte central de ese camino, tendrá que luchar con grandes gigantes (en la actualidad denominados técnicas estadísticas y algoritmos), el guía tendrá que explicarle de manera sencilla y amena, en qué consiste la lucha (las técnicas) y cómo llegar a la victoria lo más rápido posible (enseñándole a moverse por el mundo del software estadístico **R** que le permita realizar los cálculos necesarios para vencer al problema planteado a una velocidad vertiginosa).

En resumen, la información masiva y el moderno tratamiento estadístico de la misma son la

“mano invisible” que gobierna la sociedad del siglo XXI, y este manual pretende ser el guía anteriormente mencionado que le llevará de la mano cuando quiera caminar por ella.

¿Por qué este libro?

Lo dicho anteriormente ya justifica por sí solo la aparición de este manual. Sin embargo, afortunadamente, no es el primero en la materia. Son ya bastantes los materiales de calidad publicados sobre Ciencia de Datos, pero, quizás, éste pueda ser considerado el más completo. Y ello por varias razones.

La primera es su **completitud**: este manual lleva de la mano al lector desde el planteamiento del problema hasta el informe que contiene la solución al mismo; o desde no saber qué hacer con la información de la que dispone hasta ser capaz de transformar tales bases de datos masivas, y casi imposibles de manejar, en respuestas a problemas fundamentales de una empresa, institución o cualquier agente social.

La segunda es su **amplitud temática**:

- (I) Parte de las dos primeras preguntas que un neófito se puede hacer sobre esta temática: ¿qué es eso de la ciencia de datos que está en boca de todos? Y, ¿qué diablos es **R** y cómo funciona?.
- (II) Enseña cómo moverse en la jungla del *Big Data* y de los “nuevos” tipos de datos, siempre bajo el paraguas de la ética de los datos y del buen gobierno de dichos datos.
- (III) Muestra al lector cómo sacar conocimiento de la oscuridad del enorme banco de información a su disposición y que no sabe cómo abordar ni manejar.
- (IV) No deja a nadie atrás, y de forma previa al contenido central del manual (las técnicas de ciencia de datos), incluye unas breves, pero magníficas, secciones sobre los rudimentos de la probabilidad, la inferencia estadística y el muestreo, para aquéllos no familiarizados con estas cuestiones.
- (V) Aborda una treintena de técnicas de ciencia de datos en el ámbito de la modelización, análisis de datos cualitativos, discriminación, *machine learning* supervisado y no supervisado, con especial incidencia en las tareas de clasificación y clusterización -así como, en el caso no supervisado, de reducción de la dimensionalidad, escalamiento multidimensional y análisis de correspondencias-, *deep learning*, análisis de datos textuales y de redes, y, finalmente, ciencia de datos espaciales (desde las perspectivas de la geoestadística, la econometría espacial y los procesos de punto).
- (VI) Hace especial hincapié en la reproducibilidad en tiempo real (o no) entre los distintos miembros de un equipo (sea universitario, empresarial, o del tipo que sea) y en la difusión de los resultados obtenidos, enseñando al lector cómo generar informes reproducibles mediante R Markdown y documentos Quarto o en otros modernos formatos.
- (VII) Dedica un capítulo a la creación de aplicaciones web interactivas (con Shiny).

Índice general

7

- (viii) Para aquéllos con aversión a la codificación, o lo contrario, con pasión por la codificación y que quieran compartir código y colaborar con otros desarrolladores, este manual aborda la gestión rápida y eficaz de proyectos (del tamaño que sean) mediante Git, un sistema de control de versiones distribuido, gratuito y de código abierto, y GitHub, un servicio de alojamiento de repositorios Git del cual aquellos no familiarizados con la cuestión de la codificación podrán tomar el código que necesitan.
- (ix) Muestra al lector los primeros pasos para iniciarse en el geoprocесamiento en la nube.
- (x) Y, finalmente, aborda más de una docena de casos de uso (en medicina, periodismo, economía, criminología, marketing, moda, demanda de electricidad, cambio climático, predicciones de precios de vivienda, reconocimiento de patrones en la forma de tweetear...) que ilustran la puesta en práctica de todos los conocimientos anteriormente adquiridos.

La cuarta es que todo lo que el lector aprende en este manual lo puede reproducir y poner en práctica inmediatamente con **R**, puesto que el manual está trufado de *chunks* (o trozos de código **R**) que no tiene más que cortar y pegar para reproducir los ejemplos que se muestran en el libro, cuyos datos están en el paquete CDR, o utilizar dichas *chunks* para abordar el problema que le ocupa con los datos que tenga a su disposición. Una buena razón, sin duda. Por consiguiente, el manual es una buena combinación “Teoría-Práctica-Software” que permite abordar cualquier problema que el científico de datos se plante en cualquier disciplina o situación empresarial, médica, periodística...

La quinta es su **variedad de perspectivas**. Son **más de 40 los participantes** en este manual. Algunos de ellos prestigiosos profesores universitarios; otros, destacados miembros de instituciones públicas; otros, CEOs de empresas en la órbita de la Ciencia de Datos; otros, *big names* del mundo de **R** software... El manual es sin duda un magnífico ejemplo de colaboración Universidad-Empresa para buscar soluciones a los problemas de las sociedades modernas.

¿A quién va dirigido?

Fundamentos de Ciencia de Datos con R está dirigido a todos aquellos que desean desarrollar las habilidades necesarias para abordar proyectos complejos de ciencia de datos y “pensar con datos” (como lo acuñó Diane Lambert, de Google). El deseo de resolver problemas utilizando datos está en el centro de nuestro enfoque. Por tanto, como se avanzó anteriormente, este manual no deja a nadie atrás y lo único que requiere es “el deseo de resolver problemas utilizando datos”. No excluye ninguna disciplina, no excluye a las personas que no tengan un elevado nivel de análisis estadístico de datos, no excluye a nadie. Se ha procurado una combinación de rigor y sencillez, y de teoría y práctica, todo ello con sus correspondientes códigos en **R**, que satisfaga tanto a los más exigentes como a los principiantes.

También está destinado a todos aquellos que quieran sustituir la navegación por la web (la búsqueda del video, publicación de blog o tutorial *online* que solucione su problema –frustración tras frustración por la falta de consistencia, rigor e integridad de dichos materiales, así como por su sesgo hacia paquetes singulares para la implementación de las cuestiones que tratan–), por

una “**biblia de la ciencia de datos**” rigurosa pero sencilla, práctica y de aplicación inmediata sin ser ni un experto estadístico ni un experto informático.

Pero si a alguien está destinado especialmente, es a la comunidad hispano hablante. Este manual es un guiño a dicha comunidad y tiene la pretensión de que dicha comunidad tenga a su disposición, en su lengua nativa, uno de los mejores manuales de ciencia de datos de la actualidad.

El paquete CDR



El paquete CDR contiene la mayoría de conjuntos de datos utilizados en este libro que no están disponibles en otros paquetes. Para instalarlo use la función `install_github()` del paquete `remotes`.

```
# este comando sólo necesita ser ejecutado una vez
# si el paquete remotes no está instalado, descomentar para instalarlo

# install.packages("remotes")
remotes::install_github("cdr-book/CDR")
```

La lista de todos los conjuntos de datos puede obtenerse haciendo `data()`.

```
library('CDR')
data(package = "CDR")
```

Este paquete ayudará al lector a reproducir todos los ejemplos del libro. De acuerdo con las mejores prácticas en **R**, el paquete CDR sólo contiene los datos utilizados en el libro.

¿Por qué R?

R es un lenguaje de código abierto para computación estadística que se ha consolidado en las últimas dos décadas como una herramienta de primera clase para tareas de computación

Índice general

9

científica y ha sido un líder constante en la implementación de metodologías estadísticas para el análisis de datos. La utilidad de **R** para la ciencia de datos se deriva de una fantástico ecosistema de paquetes (activo y en crecimiento) y otros recursos que son excelentes para la ciencia de datos: libros, manuales, *blogs*, foros y *chats* interactivos en las redes sociales, y una gran comunidad dispuesta a colaborar, a orientar y a resolver diferentes cuestiones relacionadas con **R**.

Por otra parte, **R** es el lenguaje estadístico y de análisis de datos más utilizado en muchos entornos académicos. Y, cómo no, es utilizado por una larga lista de importantes empresas, como son Facebook (análisis de patrones de comportamientos relacionado con actualizaciones de estado e imágenes de perfil), Google (para la efectividad de la publicidad y la previsión económica), Twitter (visualización de datos y agrupación semántica), Microsoft (adquirió la empresa Revolution R), Uber (análisis estadístico), Airbnb (ciencia de datos), IBM (se unió al grupo del consorcio R), New York Times (visualización)...

La comunidad **R** también es particularmente generosa e inclusiva, y hay grupos increíbles como *R-Ladies* y *Minority R Users* diseñados para ayudar a garantizar que todos aprendan y usen las habilidades de **R**.

Agradecimientos

No queremos dar por finalizado este prefacio sin agradecer a los 44 autores participantes en esta obra su esfuerzo por condensar, en no más de 20 páginas, la teoría, práctica y tratamiento informático de la parte de la Ciencia de Datos que les fue encargada. Y no solo eso; el “más difícil todavía” fue que debían dirigirse a un abanico de potenciales lectores tan grande como personas haya con “el deseo de resolver problemas utilizando datos”. Era misión imposible. Sin embargo, a la vista del resultado, ha sido misión cumplida. El esfuerzo mereció la pena.

Además, nos gustaría agradecer el apoyo incondicional recibido por (en orden alfabético): Itzcoatl Bueno, Ismael Caballero, Emilio L. Cano, Diego Henangómez, Ricardo Pérez).

También queremos poner de manifiesto que la edición de este texto ha sido financiada por diversos entes de la Universidad de Castilla-La Mancha. En su mayor parte, por el **Máster en Data Science y Business Analytics (con R software)**, (a través de la orgánica: 02040M0280) pero también por la Facultad de Ciencias Jurídicas y Sociales de Toledo (a través de su contrato programa: 00440710), el Departamento de Economía Aplicada I (mediante sus fondos departamentales, DEAI 00421I126) y el Grupo de Investigación Economía Aplicada y Métodos Cuantitativos (que ha dedicado parte de sus fondos a la edición de esta obra).

A todos, eternamente agradecidos por ayudarnos en este reto de transformar la oscuridad en conocimiento, de convertir en una ciencia y en un arte la difícil tarea de sacar valor de los datos, el petróleo del futuro. Quizás en este momento no seamos conscientes de que hemos puesto nuestro granito de arena a la ciencia que, a buen seguro, juegue uno de los papeles más importantes de este siglo, caracterizado por el predominio de la información. Una ciencia, la Ciencia de Datos, que combina el análisis estadístico de datos, la algoritmia y el conocimiento del negocio para sacar valor del bien más abundante de la sociedad en la que vivimos: la información. Una disciplina cuyo dominio caracteriza a los científicos de datos (también denominados los

nuevos personajes del Renacimiento), profesión que ya fue calificada hace más de veinte años en la *Harvard Business Review* o en *The New York Times*, entre otros, como la “más sexy del Siglo XXI”.

Nota

Este libro está publicado por [McGraw Hill](#). Las copias físicas están disponibles en [McGraw Hill](#). La versión *online* de este libro se puede leer de forma gratuita en <https://cdr-book.github.io/> y tiene la [licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional..](#)

Si tiene algún comentario, no dude en contactar con los editores y los autores. ¡Gracias!

Capítulo 1

Ética en la ciencia de datos

Mónica Villas^a y Bilal Laouah^b

^aOdiseIA ^bAlexandria Business Solutions Based on Data

1.1. ¿Qué es la ética?

La ética es una subdisciplina de la Filosofía que estudia de manera sistemática el comportamiento humano desde las nociones del bien y del mal y en relación con la moral. Ambas disciplinas están muy relacionadas pero son diferentes: la moral tiene un carácter normativo y prescriptivo: orienta las acciones de acuerdo con algún marco de valores específico (costumbres, creencias, códigos tradicionales, normas no escritas, etc). Por el contrario, la ética está por encima de cualquier orientación particular, es decir, no se basa en ningún código de mandatos y prohibiciones concreto, sino que pretende establecer los principios a partir de los cuales evaluar las acciones y decisiones. La ética es Filosofía práctica, por eso no transmite juicios, sino que enseña a juzgar.

Las dos preguntas fundamentales que ha tratado de responder la ética a lo largo del tiempo son: ¿qué debemos hacer? y ¿qué es valioso en la vida? La primera pregunta promueve el razonamiento ético, mientras que la segunda sirve para establecer el marco de valores desde el cual juzga y razona el sujeto ético. Desde una perspectiva histórica, Aristóteles es considerado el primer autor occidental en haber sistematizado la ética. Su tratado es comúnmente conocido bajo el título de *ética a Nicómaco*.

Este capítulo se va a centrar en la ética aplicada, que es la utilización de la ética en la práctica. Algunos ejemplos de ética aplicada son la ética profesional o deontología, la bioética o la ética medioambiental. La ética aplicada a la ciencia de datos hace referencia a la reflexión que debe acompañar a la toma de decisiones en el contexto de la praxis profesional de los científicos de datos; se puede considerar, por tanto, una concreción de la ética profesional. Y es que los científicos de datos tienen que tomar decisiones a lo largo del ciclo de vida de un proyecto de datos que pueden tener consecuencias sobre las personas. Algunos procesos que pueden ser fuente de dilemas éticos son: la recopilación de datos, la transformación, la definición de objetivos a medir,

el uso de algoritmos y la explicación de los resultados. Estos pasos se pueden ver de manera detallada en el Cap.???. En todos estos pasos, el científico de datos debe usar su pensamiento crítico y tomar decisiones éticas. Así, por ejemplo, si el propósito es automatizar algún proceso, deberá reflexionar y anticipar los posibles impactos negativos que pueden derivarse ya que, si este proceso no se realiza adecuadamente, la toma de decisiones automáticas puede perpetuar algunos de los problemas éticos como son los sesgos. Para ser más específicos: supóngase que se quiere automatizar las contrataciones laborales. Para ello, el científico de datos necesita desarrollar un algoritmo que seleccione a los mejores profesionales para su compañía. Pues bien, algunas cuestiones que debe valorar son las siguientes: primero tiene que entender qué significa “los mejores profesionales” y definir los atributos que los representan. Después, tiene que buscar datos históricos de la compañía, recopilar éstos y estar seguro de que esos datos cumplen con la normativa de privacidad establecida, especialmente si la compañía reside en Europa. Aquí, el científico de datos, debe pensar en temas como la procedencia de los datos, ¿cuál es la fuente?, ¿a quién pertenecen los datos?, ¿están los datos anonimizados para que no podamos identificar a una persona de manera única?, y algunas preguntas similares referidas a la privacidad. Seguidamente, tendrá que revisar y asegurar que se tiene una muestra de datos cuyos atributos (edad, profesión, experiencia, raza, género, procedencia geográfica, etc.) no incluyen posibles sesgos, es decir, que no se tiene, por ejemplo, muchos más casos de personas de color que de raza caucásica , o más personas mayores que jóvenes, o más mujeres que hombres. En definitiva, debe asegurarse de que la muestra que tiene es suficientemente representativa de la población con la que va a trabajar y si no es así tenerlo en cuenta a la hora de analizar y comunicar los resultados. Además, se ha de tener cuidado con los datos personales, como género, edad, raza, etc., dado que, en algunos casos de uso, la toma de decisiones no debería tener en cuenta estos atributos porque podrían inducir a prácticas discriminantes, alejadas de los estándares éticos. Como se puede ver en este sencillo ejemplo, el científico de datos tiene que tomar decisiones, no sólo técnicas, que influyen en el resultado de su trabajo y que pueden afectar a otras personas. Generalmente, los científicos de datos suelen ser profesionales que provienen del mundo técnico, de carreras tecnológicas o relacionadas con las Matemáticas y, a diferencia de otros itinerarios de corte humanista, la presencia de la ética es menos frecuente, razón por la cual conviene fomentar la sensibilización respecto a estas cuestiones.

Mientras que para los profesionales de la salud existen códigos deontológicos bien establecidos y organismos que regulan la práctica de acuerdo a los mejores estándares comportamentales, no existe una guía común para el científico de datos en que se describa cómo debe comportarse. A pesar de todo, la guía de buenas prácticas que publica la asociación ACM (*Association for Computing Machinery*) puede servir de inspiración, si bien, al tratarse de meras recomendaciones, sigue siendo insuficiente para orientar éticamente el comportamiento de éstos.

1.2. Los principios éticos

Un principio no es ni más ni menos que aquello que permite preservar los derechos y libertades de las personas, sin frenar la innovación tecnológica (Olmeda and Ibáñez, 2022). La mayoría de los principios se pueden agrupar en cuatro grandes categorías: **autonomía, justicia, evitar daños y generar beneficios**. Algunos ejemplos de principios que se pueden clasificar en alguna de estas categorías son: **transparencia, explicabilidad, privacidad, accesibilidad**

1.2. Los principios éticos

13

o equidad. Aunque no hay aún un acuerdo a nivel mundial sobre cuáles deberían ser los principios claves de la IA, sí que se están desarrollando proyectos supranacionales como el de la UNESCO¹, que recientemente ha sido firmado por todos sus miembros sobre el uso ético de la inteligencia artificial.

Desde inicios del 2010, el crecimiento de la ciencia de Datos ha sido exponencial y ha comenzado a usarse en todas las industrias de manera sistemática, entre otras cosas gracias al Big Data. Actualmente, se dispone de más datos que nunca y sólo se analiza un 5 % de ellos. Además, se han producido enormes mejoras en la computación con el surgimiento de nuevos procesadores y también han ocurrido grandes cambios en el área de la algoritmia, teniendo disponibles muchos más algoritmos que nunca, lo que facilita su reutilización. Por ello, la demanda de científicos de datos que sean capaces de transformar estos datos en información clave para las empresas, ha crecido mucho en los últimos años.

Asimismo, desde 2016, distintos organismos, asociaciones, empresas y gobiernos han publicado numerosos documentos, donde se resalta la importancia de la necesidad de principios éticos para la ciencia de Datos. Google, IBM y Amazon, en el ámbito de las empresas privadas, publicaron sus principios éticos en el 2018. También son muy conocidos los principios de Asilomar de 2016 o la declaración de Toronto de 2017. La mayoría de estos documentos están desarrollados por perfiles multidisciplinares: científicos de datos, abogados o expertos en ética, que resaltan la importancia de tener en cuenta los principios éticos en la toma de decisiones automáticas cuando se utiliza la ciencia de datos.

En definitiva, las cuestiones éticas se están incorporando poco a poco en los proyectos de ciencia de datos en todo el mundo, siendo la regulación europea publicada en abril de 2021 un ejemplo a seguir. Esta regulación, diseñada a lo largo de tres años, parte de un primer documento en 2018² que ha sido liderado por un grupo de expertos de todos los países miembros, HLEGAI (*High Level Expert Group Artificial Intelligence*). A partir de este primer documento se han realizado distintas publicaciones desde 2018, que han tratado de incluir los comentarios y mejoras sugeridas por la sociedad civil, instituciones públicas, empresas e instituciones académicas, para diseñar el reciente documento de regulación de inteligencia artificial.

Para la propuesta de regulación, publicada en abril de 2021, se ha elegido un enfoque basado en riesgos, que se centra en clasificar los riesgos en cuatro tipologías:

- **Riesgo inaceptable**, como el uso de aplicaciones de *social scoring* o, por ejemplo el uso de imágenes para procesos de administración de justicia.
- **Riesgo alto**, como el uso de aplicaciones de contratación o médicas que deberán ser supervisadas por organismos designados antes de su publicación.
- **Riesgo medio**, en las que será necesaria incluir la explicabilidad necesaria para entender la toma de decisiones del algoritmo dependiendo de su tipología.
- **Riesgo bajo**, para cualquier otro tipo de aplicación.

¹(<https://en.unesco.org/artificial-intelligence/ethics>)

²<https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>

En el resto del mundo, el progreso en este tipo de regulación está siendo algo más lento, aunque países como Estados Unidos, que hasta ahora no había puesto el foco en este tipo de regulaciones, está empezando a trabajar en ello desde finales de 2021. Por otro lado, China, conocida mundialmente por su falta de respeto a la privacidad, está empezando a dar algún paso en esta área y comenzando a cambiar su política en este sentido. Como ejemplo, en marzo de 2022, se ha lanzado una regulación en la que las empresas tienen que informar mejor a los usuarios sobre sus algoritmos de recomendación. En definitiva, parece que la necesidad de la ética para proyectos de ciencia de datos está avanzando poco a poco en todas las geografías, y Europa es, por el momento, un ejemplo a seguir.

Llegados a este punto conviene distinguir entre regulaciones legales y principios éticos. Generalmente, las regulaciones legales son coercitivas y su incumplimiento puede tener consecuencias punitivas para quienes no las ratifican e implementan. Estos principios legales, para que sean legítimos, deben fundamentarse e inspirarse en ciertos valores éticos. Ahora bien, es imposible e indeseable regular legalmente todos los aspectos del comportamiento humano, de ahí la necesidad de compartir un marco de valores. La ética permitirá al científico de datos considerar cuál es la mejor decisión cuando exista un vacío legal. El razonamiento ético implica, pues, asumir la responsabilidad de pensar de manera autónoma.

Ahora bien, dado que no hay un acuerdo a nivel mundial sobre cuáles son los principios éticos más importantes para la ciencia de Datos, se han seleccionado **equidad y explicabilidad** en este capítulo. Estos principios son dos de las más relevantes a tener en cuenta por los expertos en ciencia de datos. Además, son dos de los principios en los que se centra la regulación europea.

1.3. Equidad: la importancia de los sesgos

El sesgo se puede definir como el resultado de dar un peso desproporcionado a favor o en contra de una persona o cosa en comparación con otra, y normalmente de manera injusta. El término ‘equidad’, se utiliza precisamente para tratar de que las decisiones no estén afectadas por esos sesgos. Si se analiza la literatura al respecto se pueden encontrar multitud de tipos de sesgos. En la Ciencia de Datos cuando se habla de sesgo, generalmente se hace referencia a los **sesgos algorítmicos**. Éstos, según la RAE, son “errores sistemáticos en los que se puede incurrir cuando, al hacer muestrazos o ensayos, se seleccionan o favorecen unas respuestas frente a otras”.

Este sesgo algorítmico puede suceder en cualquiera de los pasos que lleva a cabo el científico de datos como se veía en el Cap. @ref(metodología). En la Fig. 1.1, donde se representan los distintos pasos a la hora de diseñar un algoritmo de ciencia de datos, se puede ver cuáles son los momentos críticos donde, sin percibirlo, se puede caer en esos sesgos. En primer lugar, un sesgo en la **adquisición de los datos**, partiendo de muestras que ya lo tengan. En este punto se encuadran, por ejemplo, los **sesgos históricos** o los **sesgos de representación**. También puede ocurrir este sesgo en base a la selección de las características que se eligen para la construcción del modelo, que son los sesgos que se denominan **sesgos de medida**. Además, se pueden presentar sesgos en el momento del despliegue, denominados **sesgos de implementación**, que suceden cuando el contexto en el que se despliega el algoritmo es diferente del contexto en que se entrenó.

1.3. Equidad: la importancia de los sesgos

15

El estudio detallado de estos sesgos algorítmicos está enfocado a evitar que se aumenten o perpetúen sesgos de cualquier tipo, teniendo en cuenta que los algoritmos tienen como objetivo automatizar y generalizar. Como se veía en la sección anterior, mucha de la regulación que se está desarrollando en Europa va enfocada a **mantener el principio de equidad**, es decir, a tratar de evitar los sesgos en la toma de decisiones automáticas realizadas por los algoritmos que se diseñan gracias a la Ciencia de Datos.

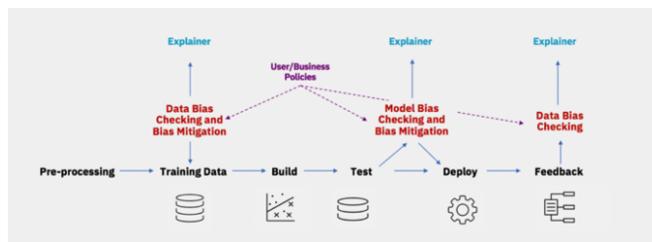


Figura 1.1: Sesgos en el proceso de machine learning. Fuente: IBM

Tomemos como ejemplo el caso de COMPAS (*Correctional Offender Management Profiling for Alternative Sanction*) es una aplicación que da soporte al sistema de justicia americana y que decide si la persona que va a ser juzgada tiene la probabilidad de ser reincidente. Este ejemplo trata de mostrar la importancia de los sesgos en el conjunto de datos, que se denominaban sesgos históricos y de representación.

En esta aplicación se establecían para cada acusado dos tipos de riesgo: **de reincidir** y **de violencia**. El índice de riesgo se establecía de 1 a 4 como bajo, de 5 a 7 como medio y de 8 a 10 como alto. Si la persona puede ser reincidente espera a que ocurra el juicio en la cárcel, y si no, no tiene que ir a la cárcel hasta que se celebre el juicio. Diversos estudios y organizaciones analizaron los datos y no parecía que hubiera ningún problema de sesgo inicialmente. Sin embargo, la organización PROPUBLICA, recogió datos de 7300 personas, durante 2013 y 2014 y demostró que la aplicación estaba sesgada.

El proceso que se siguió fue el siguiente:

1. Partiendo del **proceso de asignación de un riesgo**, se construyó el historial delictivo del acusado.
2. Para determinar la raza, se usó la clasificación establecida, de negros, blancos, hispanos y asiáticos.
3. Se **revisó la definición de reincidencia**, y cómo se establecían los riesgos en la aplicación de COMPAS.
4. Solamente se analizaron los riesgos para “**reincidencia**” y “**reincidencia violenta**”.
5. Se analizó el índice de reincidencia y de reincidencia violencia en dos años y su distribución por raza.
6. Para comprobar la disparidad entre la raza y el índice de riesgo, se utilizó una regresión logística que consideraba la raza, la edad, la historia criminal, la reincidencia futura, el grado de los cargos y el género.
7. Para ver la exactitud del algoritmo se usó un modelo de Cox.

8. Se utilizó una muestra de unos 7300 acusados (de los que se tenía datos de 2 años) para analizar la tasa de falsos positivos y falsos negativos.

Las personas de color tenían un índice de riesgo de reincidencia mucho más alto que las personas de raza caucásica. La herramienta predecía bien en el 60 % de los casos estudiados el **riesgo de reincidencia**, pero sólo en el 20 % de los casos lo hacía de manera correcta en el **riesgo de reincidir de manera violenta**. Se incluye un resumen de las conclusiones en la siguiente tabla:

Casuística en el estudio con datos de 2 años	Resultados en porcentaje
Los acusados de raza negra se les asignaba un riesgo más alto de reincidencia que los de raza caucásica	45 % a los de raza negra 23 % a los de raza caucásica
Los acusados de raza blanca se les asignaba un riesgo más bajo de reincidencia que los de raza negra	48 % a los de raza blanca 28 % a los de raza negra
Mayor asignación de riesgo de reincidencia a las personas de color	77 % más de riesgo de reincidir a las personas de raza negra que a los de raza blanca
Se determinó que las variables que tenían mayor importancia para la asignación de riesgo de reincidencia era la edad, la raza y el género	< 25 años tenía 2.5 veces más de probabilidad de ser asignado un riesgo alto 45 % si además eran de raza negra Casi un 20 % si la persona era mujer

En este caso el problema del sesgo, tiene como consecuencia que personas que no reincidirían permanezcan en la cárcel al asignarle un índice de reincidencia más alto de lo que se debería y que personas que sí podrían reincidir, quedarían en libertad por asignarles un índice más bajo de lo que sería adecuado.

Hay multitud de ejemplos publicados, respecto al tema de los sesgos, una de las mejores referencias es el libro O’neil (2016), que recopila gran variedad de casos en la que los sesgos pueden llevar a toma de decisiones erróneas y no equitativas.

1.4. ¿Es necesaria la explicabilidad?

La explicabilidad es otro de los principios clave de la propuesta europea de IA confiable y, sin duda, va a ser clave en los próximos años en cuanto la regulación europea de IA entre en vigor.

XAI (Explainable AI) es un término que acuñó DARPA (*Defense Advanced Research Project Agency*) en el año 2017 y que, agrupa dentro de explicabilidad, no sólo el concepto de interpretabilidad para los algoritmos de machine learning sino los aspectos de la Psicología que están relacionados con proporcionar explicaciones, como se puede ver en la Fig. (1.2). No se trata solamente de entender la toma de decisión del algoritmo sino de explicar por qué se toma

1.4. ¿Es necesaria la explicabilidad?

17

esa decisión de manera adecuada, dependiendo del tipo de usuarios. Si se toma el ejemplo de un algoritmo que selecciona imágenes cuando contienen un posible tumor, no serán las mismas explicaciones las que necesitará un científico de datos que un médico. Para el científico de datos será mucho más útil revisar las métricas propias del algoritmo (exactitud, precisión, sensibilidad, etc.) y, además, saber cuáles de los atributos de entrada del algoritmo han tenido más peso en la decisión. En cambio, al médico lo que le interesaría será una explicación menos técnica, más cualitativa, en la que se le explique con detalle, por ejemplo, por qué se seleccionó esa muestra frente a otras, mencionando el tamaño, la forma o características de la muestra, aspectos que van a ser entendidos por un profesional médico.

XAI: EXPLAINABLE ARTIFICIAL INTELLIGENCE

LOS CUATRO PILARES DE XAI

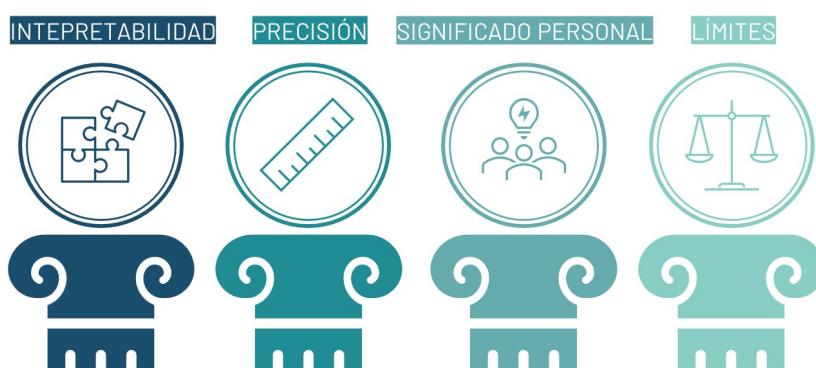


Figura 1.2: Explicabilidad según DARPA

Los algoritmos pueden clasificarse en algoritmos de “caja blanca” o transparente, (aquellos que son fácilmente interpretables), y **algoritmos opacos o de caja negra**” (aquellos que no son interpretables y que requieren de herramientas adicionales para su interpretación). Normalmente, se tiene que establecer un equilibrio entre la interpretabilidad y la exactitud, dado que son métricas que mantienen una relación inversa. A mayor exactitud, menor interpretabilidad y viceversa. Los algoritmos que son más interpretables son normalmente aquellos más sencillos, como los algoritmos de clasificación, regresión lineal o los árboles de decisión. Otros, como los modelos de *Random forest*, XGboost o algoritmos de *Deep learning*, son mucho más exactos pero no son tan interpretables y esto puede presentar ciertos problemas a la hora de usarlos en la toma de decisiones en las compañías, dado que es más difícil explicar la decisión del algoritmo. Cuando las decisiones afectan a áreas clave para las personas (decisiones médicas, de contratación, de concesión de préstamos, etc.,) es cuando es muy relevante proporcionar la explicabilidad adecuada.

Se está avanzando muy rápido en la interpretabilidad de los algoritmos, y desde 2017 se proporcionan distintas técnicas y herramientas que ayudan a ello como por ejemplo SHAP o LIME, de código abierto. En la mayoría de las ocasiones se trata de utilizar algoritmos más sencillos que ayudan a explicar otros más complejos como redes neuronales o XGboost.

Hay muchas taxonomías diferentes para la clasificación de los distintos tipos de algoritmos, una

de las más utilizadas clasifica los algoritmos en las siguientes tipologías:

- **Metodologías globales o locales:** Cuando el método utiliza una instancia para la interpretabilidad se denomina local y cuando éste usa todo el modelo se denomina global.
- **Metodologías intrínsecas o post-hoc:** ‘Intrínseca’ se refiere a cuando el método es interpretable por si mismo y post-hoc cuando es necesario usar otros algoritmos más sencillos para explicar los más complejos.
- **Metodologías ligadas al modelo o agnósticas del modelo:** Las metodologías ligadas al modelo son aquellas que se usan para un tipo de algoritmo concreto, mientras que las metodologías agnósticas permiten trabajar con cualquier tipo de modelo.

Es importante elegir la técnica más adecuada dependiendo del tipo de modelo a interpretar, así como poder combinarlas en aras de conseguir una mejor interpretabilidad. Uno de los mejores libros al respecto que recopila multitud de estas técnicas es ?.

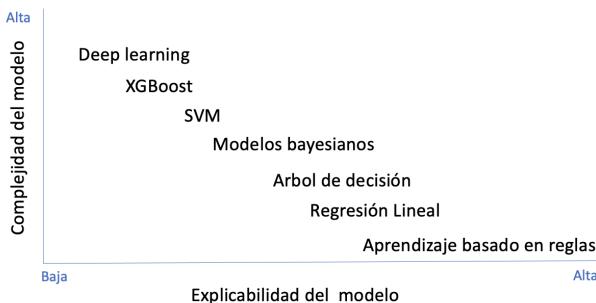


Figura 1.3: Explicabilidad vs Exactitud

1.5. Recursos en R para trabajar en sesgos y explicabilidad

Para un científico de datos es muy relevante conocer las herramientas *open source* o comerciales disponibles para que se puedan usar en análisis de sesgos o explicabilidad. Todas las herramientas en esta área son relativamente recientes. Han ido surgiendo desde el 2018 y siguen evolucionando rápidamente.

En el caso de las herramientas para detectar el sesgo, los proveedores que empiezan a incluir estos análisis son **Microsoft**, **IBM**, **Google**, **Aequitas**, **Pymetric**, **Linkedin** y el resto son *open source*. La mayoría de ellas están abiertas a contribuciones externas y todas ellas utilizan mecanismos para la detección de sesgos, aunque solamente la de Microsoft e IBM incluyen algoritmos para la mitigación de estos sesgos.

Referente a las herramientas sobre explicabilidad, los proveedores más relevantes son **Google**, **IBM**, **Oracle**, **H2O.ai**; el resto, son *open source*. La mayoría de ellas se pueden usar con

1.5. Recursos en **R** para trabajar en sesgos y explicabilidad

19

algoritmos de **caja blanca o caja negra**. Respecto a los tipos de explicaciones para distintos usuarios sólo la herramienta de IBM incluye esta funcionalidad. Se puede resaltar también, la facilidad con la que H20.ai permite elegir el nivel de exactitud y explicabilidad en el momento del diseño del algoritmo. Para un mayor detalle se puede consultar dos tablas comparativas sobre herramientas de explicabilidad y de sesgos que se incluyen en (Olmeda and Ibáñez, 2022).

Algunas de estas herramientas comerciales incluyen implementaciones en Python o **R**, de ahí la importancia de revisarlas inicialmente antes de recurrir a otro tipo de recursos.

Recursos en **R** para equidad

- Tutorial de fairness (2021)³: explica las distintas métricas usadas para medir la equidad (paridad demográfica, paridad proporcional, paridad predictiva,etc.,) y permite crear la distintas métricas y visualizarlas. El tutorial emplea los datos de COMPAS.
- Librerías de **R**⁴ incluidas en IBM fairness360 (2020): incluye algoritmos para detectar el sesgo, pero también para mitigarlo.
- Libreria EDFFair (2022)⁵: tiene una aproximación distinta, dado que permite al usuario ajustar el nivel de equidad frente a la exactitud, y poder mantener el equilibrio requerido. Esto se explica en detalle en Matloff and Zhang (2022).

Recursos en **R** para explicabilidad:

- Algunas de las herramientas más conocidas en explicabilidad que merecen mención aparte son SHAP⁶ y LIME⁷, disponibles en R y en Python y usadas en muchos paquetes comerciales. **SHAP** (2018) es uno de los métodos más usados para la explicabilidad. Utiliza los valores de Shapley para poder explicar cualquier tipo de modelo. Los detalles se pueden encontrar en Aas et al. (2021), donde también se proporciona el código⁸. **LIME** (2017) es otro de los métodos usados para la explicabilidad. Para ello, se ajusta un modelo local alrededor de un punto concreto y lo que hace es estudiar los cambios alrededor de este modelo. Se puede encontrar la explicación detallada en ?, donde, además, se proporciona el código⁹.
- El artículo Matloff and Zhang (2022) hace una recopilación de 27 librerías de R, incluyendo LIME y SHAP y el código¹⁰ puede encontrarse en github para emplear cada una de ellas.
- DALEX[`{biecek2018dalex}`] es un paquete disponible en **R** de reciente creación para ayudar a crear explicaciones partiendo de un modelo y también proporciona el código¹¹ necesario.

³<https://cran.r-project.org/web/packages/fairness/vignettes/fairness.html><https://kozodoi.me/r/fairness/packages/2020/05/01/fairness-tutorial.html>

⁴<https://developer.ibm.com/blogs/the-aif360-team-adds-compatibility-with-r/>

⁵<https://github.com/matloff/EDFFair>

⁶<https://shap.readthedocs.io/en/latest/>

⁷<https://homes.cs.washington.edu/~marcotcr/blog/lime/>

⁸https://cran.r-project.org/web/packages/shapr/vignettes/understanding_shapr.html

⁹(<https://cran.r-project.org/web/packages/lime/index.html>)

¹⁰<https://github.com/Ml2DataLab/XAI-tools/blob/master/README.md>

¹¹(<https://github.com/ModelOriented/DALEX>)

- Esta área está evolucionando mucho en los últimos años, y están surgiendo multitud de técnicas nuevas alrededor de la explicabilidad, que van a permitir entender mejor el proceso de decisión realizado por algoritmos más complejos.

Resumen

La ética, subdisciplina de la Filosofía, es el estudio sistemático del comportamiento humano desde las categorías del bien y del mal. Se trata de una rama aplicada que proporciona métodos basados en la racionalidad crítica para evaluar decisiones y acciones en base a ciertos valores compartidos. La aparición de las actuales metodologías que propone la ciencia de Datos supone un desafío que no puede resolverse únicamente a partir de criterios técnicos, puesto que muchas de las decisiones que se toman en este campo, pueden tener repercusiones sobre las personas.

Para regular ciertas prácticas, se han desarrollado diversas legislaciones en múltiples países y continentes. Estas regulaciones tienen un fundamento ético y ofrecen un marco para valorar qué acciones se ajustan a la legalidad. Ahora bien, ningún cuerpo normativo cubre todas las posibles casuísticas, razón por la cual, el razonamiento ético es fundamental para orientar la práxis de los científicos de datos.

Por otro lado, no existe un acuerdo global en relación a los principios que deben regir el comportamiento del científico de datos pero la numerosa literatura publicada desde 2016 parece ponerse de acuerdo en que todos ellos pueden agruparse en cuatro categorías: preservar la autonomía humana, generar beneficios, evitar daños y fomentar la justicia. Dado que no hay un acuerdo global a nivel mundial, este capítulo se centra en los principios clave que señala la regulación europea como son el principio de equidad y explicabilidad.

Desde 2016 han surgido herramientas que ayudan a los científicos de datos a progresar más rápido en la aplicación de los principios de equidad y explicabilidad. Algunas herramientas comerciales han ido incluyendo ciertas funcionalidades respecto a la equidad y la explicabilidad, pero además se han generado multitud de repositorios de código abierto en lenguajes como Python o R. En este capítulo se mencionan algunos de los más relevantes.

Bibliografía

- Aas, K., Jullum, M., and Løland, A. (2021). Explaining individual predictions when features are dependent: More accurate approximations to shapley values. *Artificial Intelligence*, 298:103502.
- Matloff, N. and Zhang, W. (2022). A novel regularization approach to fair ml. *arXiv preprint arXiv:2208.06557*.
- Olmeda, M. V. and Ibáñez, J. C. (2022). *Manual de ética aplicada en inteligencia artificial*. ANAYA MULTIMEDIA.
- O’neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway books.

Índice alfabético

accesibilidad, 13
algoritmo de caja negra, 17
association for computing machinery, ACM, 12
autonomía, 13

buenas prácticas, 12

deontología, 12

equidad, 13
explicabilidad, 13

filosofía, 11

impactos negativos, 12

justicia, 13

moral, 11

sesgo
 de adquisición, 14
 de implementación, 14
 de medida, 14
 de representación, 14
 histórico, 14

transparencia, 13

valor, 11

ética, 11