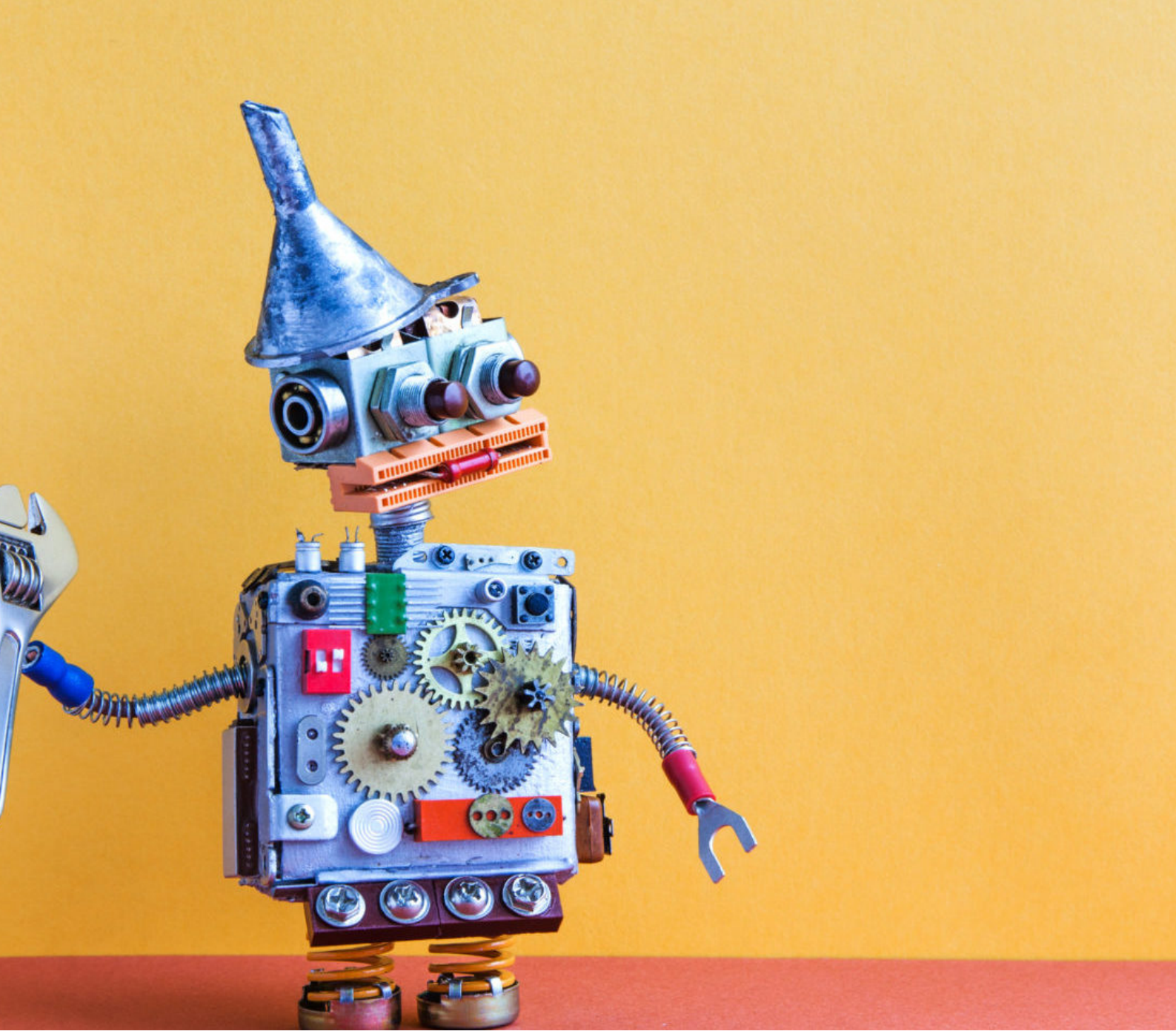


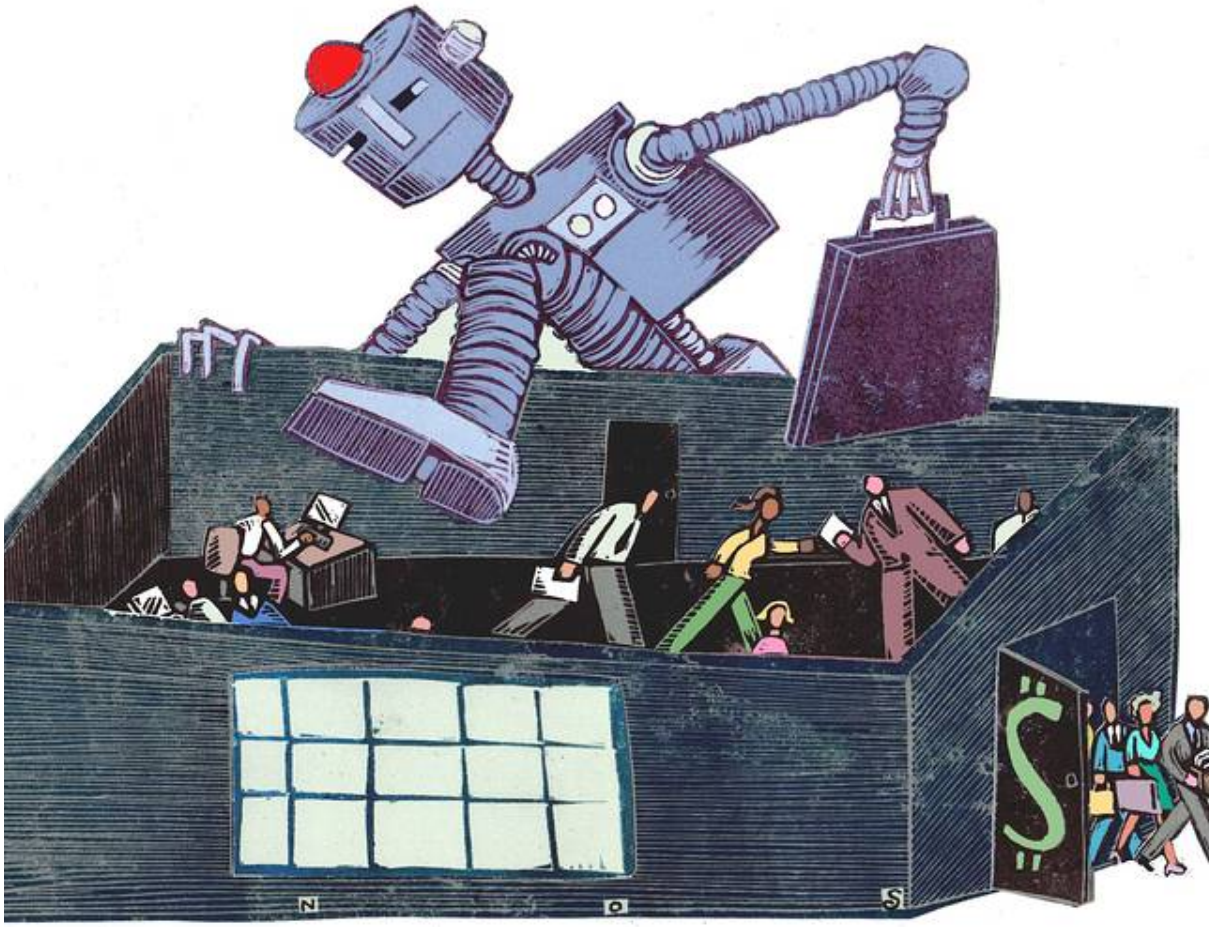
Un robot para el departamento de Recursos Humanos.

Introduciendo la Inteligencia Artificial en la gestión de los Recursos Humanos.

Clarity
Consulting
Group

Un robot para el departamento de Recursos Humanos.

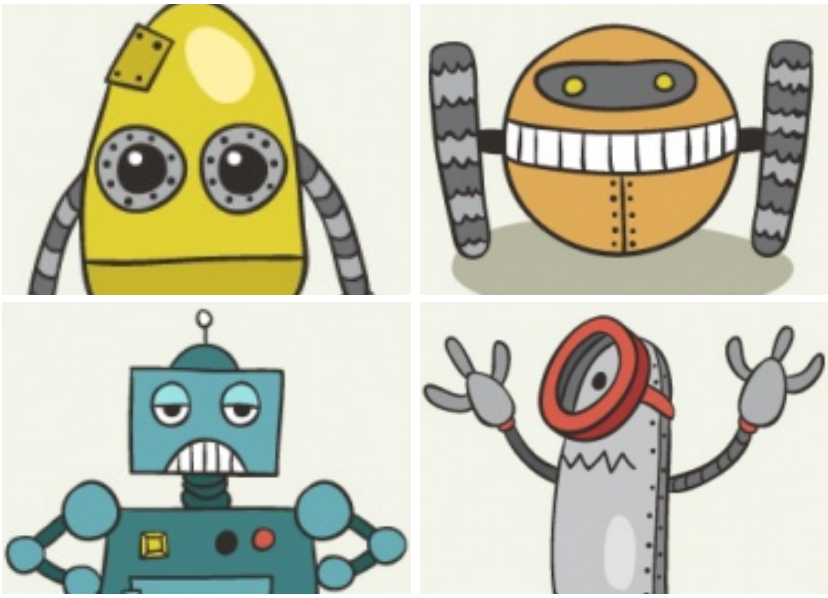
Introduciendo la Inteligencia Artificial en la gestión de los RRHH.



La velocidad con la que la retórica sobre la transformación digital ha pasado del"Big data" (BD) al "Machine Learning" (ML) y después a la inteligencia artificial (IA) es asombrosa. Sin embargo, la correspondencia entre la retórica y la realidad es una cuestión diferente. La realidad es que la mayoría de las compañías están todavía luchando para lograr algún progreso en la construcción de capacidades de análisis de datos: según IBM, el 41% de los CEOs afirman que no están preparados para hacer uso de las nuevas herramientas de análisis de datos y sólo el 4% afirma que están preparados "en gran medida" (IBM 2018).

En los últimos años se han producido importantes avances en los campos del reconocimiento de patrones, el procesamiento del lenguaje natural (PNL) y el aprendizaje profundo mediante redes neuronales. Estos avances, nos han acercado a la IA que

“Inteligencia artificial debe ser inteligencia aumentada. De lo contrario continuar con nuestros sesgos actuales seguirá mereciendo la pena”.



representa la capacidad de las máquinas para imitar la toma de decisiones humana de forma adaptativa.

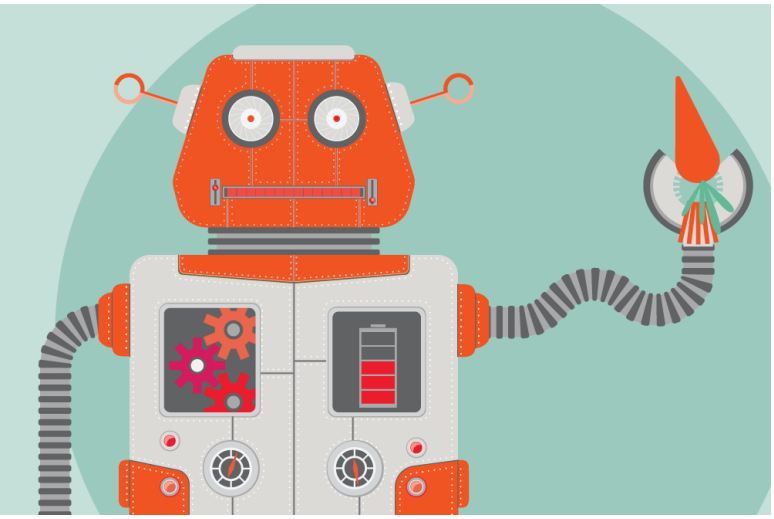
Sin embargo, con respecto a la gestión de los recursos humanos, donde la promesa de decisiones más sofisticadas se ha desvelado como una verdadera necesidad en los últimos tiempos, la realidad es que pocas organizaciones han entrado incluso en la primera gran fase de recopilación y construcción de una base de datos operativa. Sólo el 22% de las empresas afirman haber adoptado algún tipo de herramienta analítica en recursos humanos (LinkedIn 2018).

Los retos de la aplicación de la IA a la gestión de RRHH.

La analítica de datos (Big data) es probablemente más fácil de ver en campos como el marketing. Si bien en el área de marketing hay muchas preguntas que todavía deben encontrar respuesta, éstas tienden a distinguirse por su relativa claridad. Por ejemplo, la predicción acerca de quién comprará un producto o cómo los cambios en su presentación afectan sus ventas. Además, los resultados se miden fácilmente y a menudo se recogen electrónicamente en el proceso de venta. Por si esto fuera poco, el número de observaciones es muy grande (ventas de un determinado artículo en todo el país a lo largo del tiempo, por ejemplo), lo que hace factible la aplicación técnicas de análisis de grandes bases datos.

Por otro lado, la cuestión ética es más ligera en el ámbito del marketing que en el de RRHH. Aunque el marketing no está exento de problemas éticos, la idea de que las empresas deben esforzarse en vender más cantidad de sus productos está bien aceptada en general, al igual que la idea de que las empresas tratarán de influir en los clientes para que compren más.

Sin embargo, la aplicación efectiva de la IA a los problemas de recursos humanos presenta desafíos muy diferentes. La construcción de algoritmos o sistemas de IA para la ayuda a la toma de decisiones o para predecir futuros comportamientos de los empleados todavía está muy lejos de ser una realidad operativa.

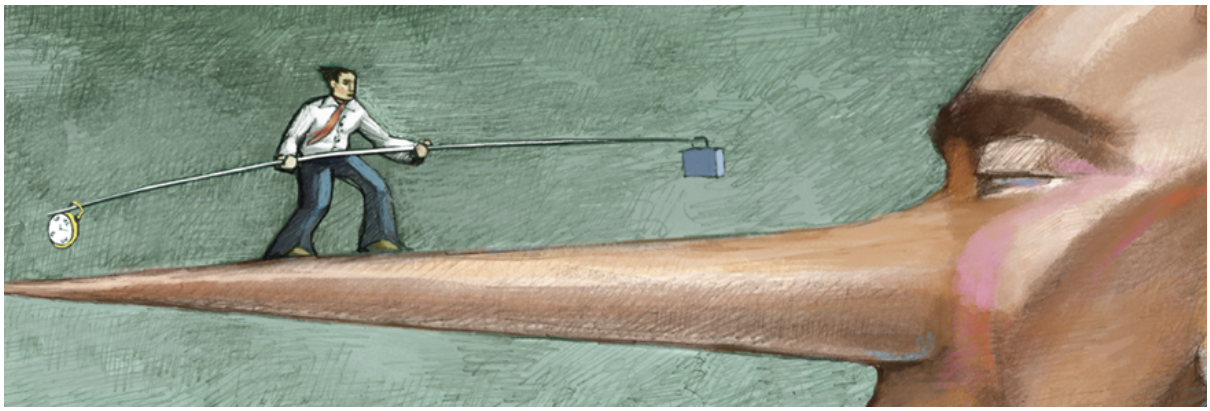


Un primer problema es la complejidad de los procesos que gestiona la función de RRHH. Definir, por ejemplo qué es ser un buen empleado no es una tarea fácil. Hay muchas dimensiones en ese constructo y medirlo con precisión, para la mayoría de los puestos, es bastante complejo. Variables como las puntuaciones de la evaluación del desempeño (la métrica más utilizada) han sido criticadas rotundamente por problemas de validez y fiabilidad, así como por sesgos emocionales y cognitivos y algunas compañías están renunciando a ellas por completo (Cappelli y Tavis 2017). El hecho es que cualquier puesto de trabajo razonablemente complejo es altamente interdependiente con otros puestos y por lo tanto el desempeño individual es difícil de separar del desempeño grupal (Pfeffer y Sutton 2006).

En segundo lugar, los conjuntos de datos sobre recursos humanos tienden a ser bastante pequeños para los estándares del Big Data. El número de empleados que incluso una gran empresa puede tener es trivial comparado con el número de compras que hacen sus clientes. Las técnicas de la ciencia de datos funcionan mal cuando predicen resultados relativamente raros (estadísticamente hablando)

En tercer lugar y posiblemente lo más importante, los resultados de las decisiones sobre recursos humanos, como por ejemplo quién es contratado y quién es despedido, tienen consecuencias tan graves para las personas y la sociedad que es necesario elaborar marcos jurídicos que rijan la forma en que los empleadores deben tomar esas decisiones, planteando cuestiones de justicia y equidad. La sociedad en general vigila los resultados de las actividades de las empresas con sus recursos humanos. Las decisiones de empleo están sujetas a una serie de complejas preocupaciones socio-psicológicas, tales como el valor y el estatus personal, la imparcialidad percibida, las expectativas contractuales y relacionales, que afectan tanto a los resultados de la organización como a los individuales. En consecuencia, poder explicar y justificar el modelo de preferencia y el funcionamiento de los algoritmos ante este tipos de decisiones es mucho más importante en el área de RRHH que en otros campos.

Por último, las reacciones de los empleados pueden ser motivo de preocupación. Ya que los trabajadores son participantes activos en las operaciones de una empresa, existe la posibilidad de que sean capaces de eludir sistema o de reaccionar de forma adversa a las decisiones basadas en algoritmos.



Para ilustrar estas preocupaciones, vamos a considerar el uso de un algoritmo para predecir el mejor candidato en un proceso de selección basado en los atributos de los "buenos empleados" en la plantilla actual.

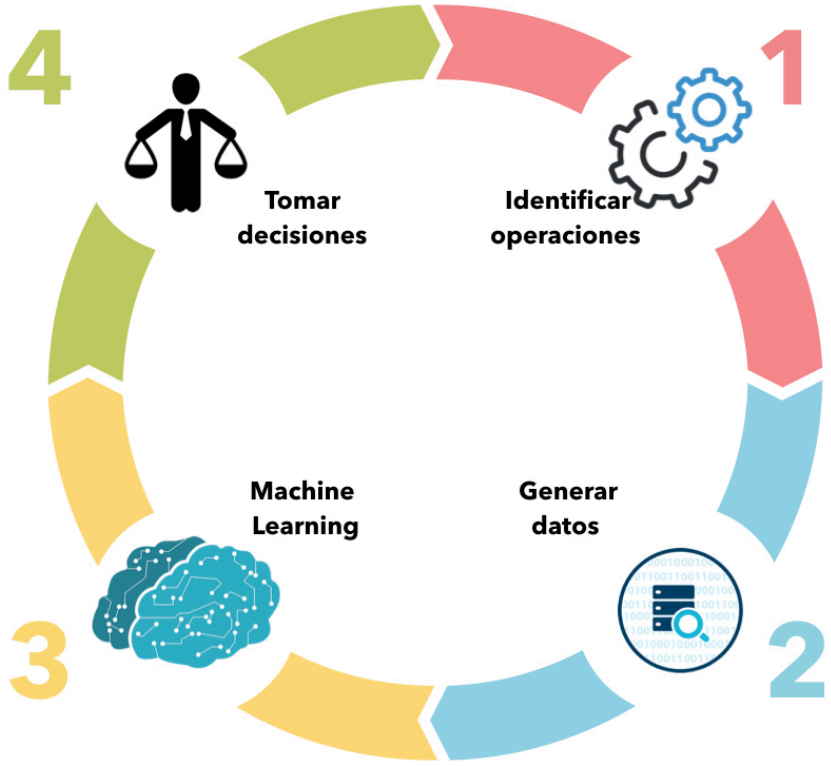
Incluso si pudiéramos demostrar una relación causal entre el sexo y el rendimiento laboral, es posible que no confiemos en un algoritmo que diga que resulta necesario contratar a más hombres que mujeres. Hay varias razones: el desempeño laboral en sí mismo puede ser un indicador sesgado, los atributos de la fuerza laboral actual pueden estar distorsionados por la forma en la que hemos contratado en el pasado (por ejemplo, contratamos pocas mujeres) y tanto el sistema legal como las normas sociales nos crearían problemas sustanciales si actuáramos de acuerdo con él.

Si, en otro ámbito, pretendemos construir un algoritmo para predecir comportamientos basados en los despidos por mal rendimiento, el número de tales casos en una compañía típica es demasiado pequeño para construir un algoritmo efectivo.

Además, una vez que los solicitantes descubren el contenido de nuestro algoritmo de contratación, es probable que respondan de manera diferente en las entrevistas y hagan que el algoritmo carezca de valor.

Un modelo de desarrollo para el área de IA en RRHH.

El proceso de implantación de un sistema de IA para el área de RRHH debería discurrir a lo largo de un proceso con las siguientes 4 fases: Identificar las operaciones sobre las que vamos a aplicar la IA, generación de datos, aprendizaje automático (Machine Learning) y toma de decisiones.



1. Identificación de operaciones de RRHH.

Una de las razones del interés en aplicar las herramientas de la Inteligencia artificial y la ciencia predictiva a los recursos humanos es que implican un buen número de operaciones en las que están en juego altos costes para la empresa.

A continuación se presentan algunos ejemplos de las operaciones más comunes en recursos humanos junto con las tareas de predicción correspondientes:

Proceso de RRHH	Objeto de la de predicción
Reclutamiento: identificar a los posibles candidatos y persuadirlos para que presenten su candidatura	¿Estamos asegurando nos llegan buenos candidatos?
Selección: elegir qué candidato debe recibir la oferta de trabajo.	¿Estamos ofreciendo trabajo a los que serán los mejores empleados?
"On boarding": incorporación de un empleado a una organización	¿Qué prácticas hacen que las nuevas contrataciones entren en productividad más rápidamente?

Proceso de RRHH	Objeto de la de predicción
Formación y mejora del rendimiento.	¿Qué intervenciones tienen sentido y para qué individuos? ¿Mejoran realmente el rendimiento?
Gestión del rendimiento: identificar el buen y el mal rendimiento	¿Mejoran nuestras prácticas el rendimiento laboral?
Talento: determinar quién es promovido	¿Podemos predecir quién tendrá más éxito en nuevos puestos?
Retención.	¿Podemos predecir quién es probable que se vaya y gestionar el nivel de retención?
Prestaciones a los empleados.	¿Podemos identificar qué beneficios son los más importantes para que los empleados sepan qué elegir cuando hay opciones y cuáles son los efectos de esos beneficios (por ejemplo, si mejoran el reclutamiento y la retención)?

2. Generación de datos.

Cada una de estas operaciones implica tareas administrativas, cada una afecta al desempeño de la organización de manera importante y cada una incluye puestos específicos, roles de trabajo, procedimientos, así como interacciones con otras áreas o departamentos. Estas operaciones producen volúmenes de datos en forma de textos, archivos, ficheros informáticos etc… A medida que las operaciones se trasladan al espacio virtual, muchos de estos productos se presentan en forma de "emboscada digital", es decir, datos de rastreo sobre actividades digitales (por ejemplo, solicitudes de empleo online) que pueden utilizarse para crear algoritmos de reclutamiento.

Los sistemas de información de recursos humanos, los sistemas de seguimiento de los candidatos, “emboscada digital” y otros marcadores son entradas críticas para la etapa de "generación de datos". Típicamente, estas entradas tiene que ser extraídas de múltiples bases de datos, convertidas a un formato común y unidas antes de que el análisis pueda llevarse a cabo. Todo esto presenta múltiples problemáticas que abordaremos posteriormente

3. Aprendizaje automático (Machine Learning)

Por "aprendizaje automático" (ML) nos referimos a un amplio conjunto de técnicas que pueden adaptarse y aprender de los datos para rendir cada vez mejor en una tarea; la tarea en cuestión es la predicción. Un científico de datos elige un algoritmo de

aprendizaje, determina la métrica más apropiada para evaluar su precisión y entrena el algoritmo utilizando la muestra de entrenamiento.

Algunos de los algoritmos de predicción más utilizados, como la regresión logística y las redes neuronales, infieren la variable que se quiere predecir a partir de correlaciones estadísticas entre otras variables observadas. La precisión de los modelos preliminares se evalúa en los datos de desarrollo hasta que se estabiliza a un nivel aceptable.

Para la contratación, por ejemplo, podríamos ver qué características de los candidatos se han asociado con un mejor desempeño laboral y utilizarlas para seleccionar candidatos en el futuro.

La Inteligencia Artificial es solo una ayuda. Las decisiones las seguirán tomando los gestores.

Pero también podemos hacer análisis y definir algoritmos para hacer recomendaciones a los empleados acerca de las trayectorias de formación que deben iniciar o sobre cuáles son los puestos que tienen más sentido para ellos.

Todo esto, de la misma manera que Netflix puede recomendarnos contenidos basados en nuestras preferencias o Amazon puede recomendar productos basados en el comportamiento de compra o navegación o Spotify nos recomienda canciones basándose en nuestra historia de búsqueda de artistas.

Estos algoritmos difieren en algunos aspectos importantes de los enfoques tradicionales utilizados en RRHH. En la psicología industrial, el campo que históricamente ha centrado la mayor atención es la contratación y la investigación se ha centrado en probar hipótesis explicativas acerca de la relación entre los predictores individuales y el rendimiento laboral.

El aprendizaje automático, en cambio, genera un algoritmo que puede hacer uso de muchas variables, que pueden no estar en el cañón de la literatura teórica asociada al tema. De hecho, uno de los atractivos del aprendizaje automático es investigar factores no tradicionales porque el objetivo es construir una mejor predicción, no avanzar en la teoría del campo en el que se basa el investigador.

4. Toma de decisiones.

La etapa final, trata de la forma en que utilizamos los conocimientos del modelo de aprendizaje de la máquina (ML). No se trata de estandarizar las decisiones de Capital Humano. El escenario ideal es aquel en el que el gestor de RRHH tienen la opción de ignorar la evidencia sobre las predicciones o usarlas como mejor le parezca,

integrándolas dentro del mundo de sus propios valores y juicios para generar una mejor toma de decisiones. Hablamos de "Inteligencia Aumentada".

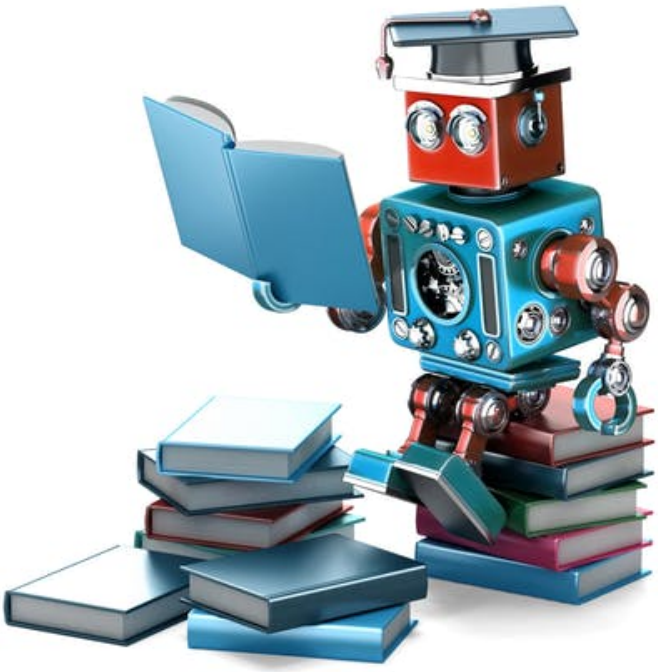
Y entonces, ¿cómo avanzamos?

Afrontando los retos de la IA para RRHH.

En esta sección, exploramos en detalle los cuatro desafíos generales de la IA que se describen en la Introducción: la complejidad del fenómeno de los RRHH, la escasez de datos, las limitaciones éticas y legales y las reacciones de los empleados a la utilización de la IA para las decisiones de RRHH.

Para hacer que estos desafíos sean manejables, los discutimos en el contexto de las etapas del Ciclo de Vida de la IA en las que son más relevantes.

Etapa de generación de datos.



La complejidad inherente a muchos fenómenos de RRHH se manifiesta en la fase de generación de datos. La fuente más importante de complejidad puede ser el hecho de que no es fácil medir, por ejemplo, lo que constituye un "buen empleado", dado que los requisitos del trabajo son amplios, el seguimiento de los resultados del trabajo es deficiente y los sesgos asociados con la evaluación del rendimiento individual son numerosos. Además, los trabajos complejos son interdependientes entre sí y por lo tanto, el rendimiento de un empleado es a menudo inextricable del rendimiento del grupo. Sin una definición clara de lo que significa ser un buen empleado, un gran número de operaciones de RRHH se enfrentan a dificultades considerables.

En cuanto a los datos, no todos los detalles de las operaciones dejan huellas digitales y no todas las huellas que quedan pueden extraerse y convertirse en un formato utilizable a un coste razonable. Por ejemplo, responsables de selección podrían rastrear los canales a

través de los cuales los candidatos acuden a ellos - desde referencias personales, visitas a nuestro sitio web, bolsas de trabajo y así sucesivamente. La mayoría de las empresas no retienen los datos de los candidatos que seleccionan. Estas opciones limitan los tipos de análisis que se pueden realizar y las conclusiones que se pueden sacar.

Por otro lado, estos datos podrían ser compartidos entre diferentes empresas de manera anónima, aumentando con ello la dimensión de la base de datos. Sin embargo, no existe una lista de variables "estándar" que los empleadores decidan recopilar y retener a través de sus operaciones de RRHH. Esto reduce la medida en que las mejores prácticas de análisis pueden transferirse de una organización a otra. Las medidas conductuales de las encuestas de clima o de compromiso, por ejemplo, varían considerablemente en su uso en todas las organizaciones y además las diferencias en la contabilidad de costes provocan que los detalles que las compañías tienen sobre los costes de las diferentes operaciones de RRHH difieran enormemente (por ejemplo, cómo se detallan los costes de formación en la contabilidad ¿qué nivel de desagregación tienen estos datos?.

La complejidad de los fenómenos de RRHH en la Etapa de Generación de Datos no es del todo nueva. Al abordar estos desafíos, los empleadores pueden beneficiarse de las lecciones clave de la gestión del rendimiento:

- No busque medidas perfectas de rendimiento, que de todos modos no existen. Es mejor elegir medidas razonables y atenerse a ellas que seguir dando vueltas a los sistemas para encontrar la medida perfecta.

- Agregar información desde múltiples perspectivas y a lo largo del tiempo. Las herramientas digitales de RRHH permiten evaluaciones rápidas y en tiempo real entre colegas que utilizan dispositivos móviles.

- Utilizar medidas objetivas de los resultados del desempeño basadas en objetivos y KPIs determinados ex ante siempre que sea posible, pero complementarlos con evaluaciones más subjetivas para captar resultados menos tangibles, como por ejemplo si el empleado encaja en la cultura deseada de la empresa.

- Integrar los datos de RRHH con los datos financieros y de negocio de la empresa para analizar los efectos en el rendimiento de la unidad de negocio.

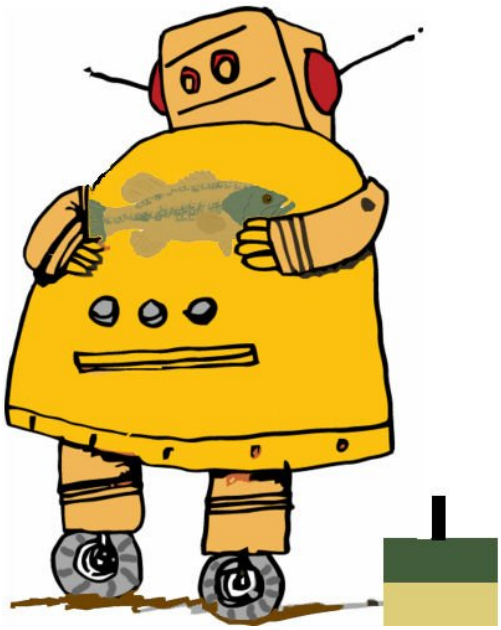
La complejidad de los fenómenos de RRHH encuentra otro problema en forma de la variedad de proveedores especializados que se ocupan de una sola tarea. Es muy común que un empleador utilice a un proveedor para realizar el proceso de evaluación del rendimiento que genera las puntuaciones de rendimiento de los empleados, de otro para el software de seguimiento de candidatos en los procesos de selección, de un tercero para los datos de compensación y nómina, y así sucesivamente. Por tanto, la agregación de los datos para el análisis se convierte en un enorme desafío.

Para ilustrar lo rudimentario que siguen siendo la mayoría de los esfuerzos de gestión de bases de datos existentes en las operaciones de RRHH, la gran mayoría de los gestores de

RRHH nos dicen que el software que utilizan con más frecuencia para organizar y gestionar sus datos es Excel. Muy pocos utilizaron herramientas más específicas y potentes como "**Tableau**".

Antes de lanzar un gran proyecto de RRHH digital, las empresas deben determinar qué datos son necesarios y auditar lo que está disponible y puede ser extraído y transferido a un formato utilizable con costes razonables. Por ejemplo, si el empleador desea utilizar un algoritmo de aprendizaje automático en la contratación, necesita recopilar datos históricos de aquellos candidatos que "no fueron contratados".

Para demostrar su compromiso con la transformación digital, así como para beneficiarse de ella, la alta dirección de la empresa tiene que hacer del intercambio de datos una prioridad a corto plazo e invertir en la estandarización de datos y la integración de plataformas a largo plazo.



Dadas todas estas preocupaciones sobre la base de datos, puede ser costoso tratar de responder con IA una pregunta de RRHH por primera vez. Los responsables del análisis de datos, por lo tanto, tienen que tener cuidado sobre dónde "apostar" en términos de reunir datos para realizar aquellos análisis que permitan arrojar resultados verdaderamente aplicables.

Small Data vs Big Data.

Small Data es una pieza fundamental para el análisis de recursos humanos. El Small Data es un conjunto de datos con un formato y un volumen que los hace accesibles, informativos y procesables. Por tanto, el objetivo del Small Data es hacer entendible el Big Data, conectando, organizando y empaquetando los datos para que estén disponibles para todos los miembros de la organización, siendo fáciles de aplicar en su día a día, y se centren en la tarea en cuestión que se desea abordar.

La mayoría de las empresas no contratan a muchos trabajadores en un año, ni hacen suficientes evaluaciones de desempeño o recolectan suficientes puntos de datos para que su fuerza laboral actual utilice técnicas de aprendizaje automático porque no tienen tantos empleados. Sin embargo, la literatura de aprendizaje automático ha demostrado que el acceso a datos más grandes tiene ventajas sustanciales en términos de precisión predictiva (Fortuny, Martens y Provost 2014).

Al mismo tiempo, los “datos pequeños” a menudo son suficientes para identificar las relaciones causales, que los gestores necesitan entender para poder actuar. La literatura de gestión tiene una ventaja importante sobre la ciencia de los datos en la articulación de las relaciones causales.

Cuantos menos datos tengamos, más teoría necesitaremos para identificar los predictores causales. La teoría también puede provenir del conocimiento experto y de la experiencia de gestión. Sin embargo, la IA requiere que los gestores pongan sus suposiciones sobre la mesa y persuadan a las otras partes interesadas de su exactitud. La formulación de tales suposiciones (y valores) a menudo se convierte en una competencia entre las partes interesadas. Cuando un proceso formal revela grandes desacuerdos, una alternativa podría incluir la recopilación de datos adicionales de experimentos aleatorios con el fin de probar los supuestos causales. Google se hizo famoso por realizar experimentos para todo tipo de fenómenos de RRHH, desde el número óptimo de entrevistas por candidato hasta el tamaño óptimo del plato de comida en la cafetería (Bock 2015). Si las discusiones, los experimentos y la persuasión de los líderes no conducen a un consenso razonable sobre el modelo causal que genera el resultado de interés, es probable que los análisis de IA sean contraproducentes y, por lo tanto, deben evitarse hasta que se puedan recopilar más o mejores datos.

Sumando esfuerzos en la recopilación de datos.

Uno de los atractivos de utilizar a los proveedores es su capacidad de combinar los datos de muchos empleadores para generar sus algoritmos. Estos enfoques ya se han utilizado en las pruebas de selección estándar o como a veces se les conoce ahora, en las pruebas previas al empleo, como las que se realizan para los puestos de ventas. Por ejemplo, la conocida empresa ADP, que maneja operaciones de externalización de nóminas, ha podido aprovechar esta escala para construir modelos predictivos de compensación y rotación de plantilla. Las empresas clientes están dispuestas a poner sus datos a su

No busque medidas de rendimiento perfectas, simplemente no existen.

disposición para este ejercicio a cambio de tener acceso a comparaciones con el benchmark.

La complicación para los empleadores individuales es saber hasta qué punto su contexto es lo suficientemente distinto como para que un algoritmo basado en datos de otros lugares genere predicciones efectivas en su propia organización.

Reacciones de los empleados a los esfuerzos de recolección de datos.

El hecho de que los empleados puedan sesgar sus respuestas y los datos en función de cómo piensan que se utilizarán los datos es una preocupación importante para todos los esfuerzos analíticos de RRHH. Debido a esto, existe la preocupación de encontrar fuentes alternativas de datos que puedan considerarse más fiables.

Las empresas están empezando a utilizar ahora la información de las redes sociales para la contratación (por ejemplo, buscando pruebas de “mal comportamiento”, buscando pruebas de aptitud etc…); otros lo utilizan para evaluar el "riesgo de fuga" o los problemas de retención (por ejemplo, la identificación de los perfiles actualizados de LinkedIn).

Otra práctica emergente, aunque no exenta de polémicas el análisis de los datos de los correos electrónicos en busca de evidencias para la prevención de malas prácticas. Por ejemplo, la aparición de términos como "acoso" en el tráfico de correo electrónico puede desencadenar una investigación interna para detectar problemas en el lugar de trabajo. Sin embargo, una vez que los empleados entienden estos procesos, adaptarán su lenguaje para ayudar a generar investigaciones o para prevenirlas.

Otra práctica emergente es el procesamiento de lenguaje natural para medir el tono de los comentarios que los empleados publican en los foros de chat internos o externos, ayudando así a predecir el riesgo de fuga de los empleados.

Ante estas prácticas, es fácil sentir que hay un problema ético en el análisis de los datos generados en las redes sociales o en el correo electrónico. Y las compañías deben tomar una posición antes de empezar y actuar en consecuencia. hay quienes piensan que el uso de estos datos (tanto públicos, como existentes en la herramientas y redes internas) es aceptable. Otros, consideran que algunas de las fuentes de datos (como los generados mediante el análisis de correo electrónico están fuera de los límites y otros piensan que cualquier dato relacionado con los empleados es apropiado para su uso, siempre y cuando estén “anonimizados”.

Muchas de estas y otras consideraciones similares entran en el ámbito de la privacidad, que adquiere nuevas dimensiones en la era digital. Los datos pueden persistir mucho más allá del tiempo en que fueron generados y las compañías pueden utilizarlos para fines no previstos por el creador, por ejemplo, las palabras utilizadas en un intercambio de correo electrónico con un colega pueden utilizarse para predecir el riesgo de fuga. También puede afectar inadvertidamente a otras personas, por ejemplo, a los amigos de un empleado que hayan sido etiquetados en fotos en redes sociales.

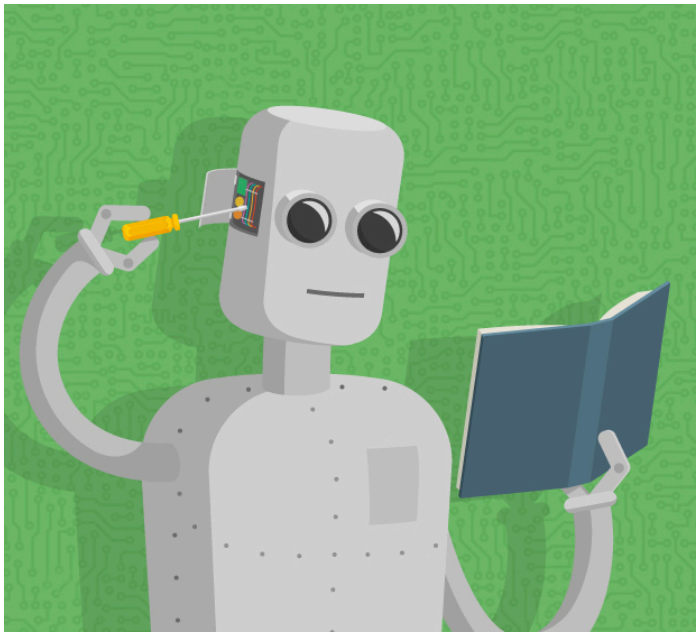
En este caso, los empleadores deben tener en cuenta las regulaciones en materia de privacidad, como "el derecho a ser olvidado" o el Reglamento General de Protección de Datos (GDPR) de la UE. El primero afirma que las empresas tienen que satisfacer las demandas de los particulares de eliminar sus trazas digitales después de un cierto período de tiempo; el segundo es un tratamiento completo de todos los aspectos de la privacidad de los datos en la era digital (www.eugdpr.org).

En cuanto a las soluciones tecnológicas a la cuestión de la privacidad de los datos, los informáticos están trabajando activamente en métodos analíticos de preservación de la privacidad de los datos que se basan en la noción de privacidad diferencial en la construcción de algoritmos. Se trata de que los datos son aleatorizados durante el proceso de recolección, lo que lleva a "no aprender nada sobre un individuo mientras se aprende información útil sobre la población". Los analistas no saben qué datos se utilizan en el análisis y qué datos se sustituyen por ruido, pero sí conocen el procedimiento de generación de ruido y, por lo tanto, pueden estimar el modelo de todos modos.

Etapa de aprendizaje automático

Puede que no resulte sorprendente que un algoritmo de LD para decidir qué candidatos contratar arroje mejores resultados que cualquier otra cosa que una empresa haya usado antes (incluyendo una bola de cristal). De hecho, una queja razonable de los gestores de RRHH es que, hasta ahora, la investigación en recursos humanos no ha sido muy eficaz a la hora de ayudar a las empresas.

De hecho, la mayoría de los predictores del rendimiento laboral que se han usado hasta ahora, como la personalidad y las pruebas de CI o de actitudes, predicen tan poco del



desempeño en el trabajo que se está creando una enorme oportunidad para el análisis de datos.

Por ejemplo, la empresa HireVue ayuda a sus clientes a seleccionar personal mediante entrevistas en vídeo. Parte de su oferta incluye algoritmos que analizan las señales verbales y no verbales de un candidato que son capturadas en video para predecir el futuro desempeño del candidato en la compañía. Estos algoritmos son entrenados constantemente con los datos de todos los clientes de Hirevue en base a aquellos que han demostrado tener éxito en sus puestos después de la contratación.

Un problema potencial derivado del uso de “los mejores” para entrenar algoritmos es que introduce un sesgo en la variable dependiente, al examinar sólo a aquellos que tienen éxito. Por lo tanto, no podemos identificar lo que distingue a los mejores de los demás.

Varios aspectos del proceso de definición del modelo también pueden ser desafiantes. Por ejemplo, puede haber más de una medida de "ajuste" con los datos. Cuando consideramos la diferencia entre las poblaciones mayoritarias (por ejemplo, los empleados hombres) y las poblaciones minoritarias (por ejemplo, los empleados mujeres), los algoritmos que maximizan el éxito predictivo para la población en su conjunto pueden discriminar el éxito predictivo para la población minoritaria. La generación de algoritmos separados para cada uno de ellos, que podrían conducir a mejores resultados, entra en conflicto con las normas legales y éticas de un tratamiento dispar. Como resultado, existen equilibrios fundamentales en este campo entre la exactitud y la imparcialidad que deben ser valorados en cualquier implementación de Machine Learning.

Los sesgos introducidos en los algoritmos.

Un caso bien conocido a la hora de definir la imparcialidad en los algoritmos de aprendizaje automático es el de la investigación de ProPublica sobre el uso de software de aprendizaje automático por parte de jueces en el condado de Broward (Florida) para determinar si una persona acusada de un delito debe ser puesta en libertad bajo fianza. Se argumentó que la herramienta que utilizaban estaba sesgada en contra de los acusados negros. Surgió un debate entre ProPublica y los desarrolladores del software sobre la cuestión de si el algoritmo estaba o no sesgado. Gran parte de las diferencias

entre las dos partes se deben en última instancia al hecho de que existen múltiples maneras de definir la equidad en el proceso de toma de decisiones. Por ejemplo, un punto de vista de imparcialidad podría implicar la exclusión total de la variable de raza del algoritmo, mientras que otro podría centrarse simplemente en maximizar la probabilidad de identificar a los acusados, independientemente de su raza, que reincidan en el delito. Otro más podría centrarse en asegurar que la frecuencia de clasificar erróneamente a los acusados como reincidentes no sea mucho mayor para una raza que para otra. Por lo tanto, a la hora de desarrollar el algoritmo, puede ser difícil ponerse de acuerdo en los criterios discriminantes que tienen implicaciones fundamentales para el diseño.

Etapa de toma de decisiones

Existen tres desafíos principales cuando los responsables de la toma de decisiones tratan de aplicar las predicciones producidas por el aprendizaje automático. El primero se refiere a la imparcialidad y a cuestiones legales, el segundo se refiere a la falta de explicabilidad del algoritmo, y el tercero a la cuestión de cómo reaccionarán los empleados ante las decisiones algorítmicas.

Imparcialidad

En el contexto de los recursos humanos, existen numerosas cuestiones relacionadas con la equidad. Uno de los más obvios es el reconocimiento de que es muy probable que cualquier algoritmo mire hacia atrás. La presencia de algún tipo de discriminación pasada en los datos utilizados para construir un algoritmo de contratación, es probable que conduzca a un modelo que pueda seleccionar desproporcionadamente a aquellos perfiles que fueron contratados en el pasado. Al utilizar estos algoritmos se corre, por tanto, el riesgo de reproducir la diversidad demográfica (o la falta de ella) que existe en los datos históricos.

En el contexto de los recursos humanos, existe ya una creencia generalizada de que las evaluaciones de los candidatos y empleados están fuertemente influenciadas por los sesgos (prejuicios) del evaluador, más comúnmente relacionados con la demografía. Los algoritmos pueden reducir ese sesgo al estandarizar la aplicación de criterios a los resultados y al eliminar información que es irrelevante para el desempeño, pero que podría influir en las decisiones de contratación.

Sin embargo, los factores que pueden ser importantes para dotar de poder predictivo al algoritmo pueden ser considerados también inapropiados, como el estatus social. ¿Nos llevará eso a eliminar esos criterios y a perder poder de predicción?

La posibilidad de que se presenten impugnaciones judiciales, especialmente en relación con el impacto adverso de las decisiones en materia de empleo, suscita diferentes preocupaciones. Por ejemplo, dejar que los supervisores y jefes de equipo tomen decisiones de empleo sin orientación probablemente conducirá a un sesgo mucho mayor

que el que generan los algoritmos. Pero ese sesgo es mucho más difícil de responsabilizar porque no es sistemático y se limita a ese grupo de trabajo. Los algoritmos utilizados en toda la organización pueden tener menos sesgo que confiar en los criterios de supervisores dispares, pero el sesgo resultante es más fácil de identificar y tiene efectos sobre clases enteras de individuos, llevando a potenciales demandas. ¿Valdrá la pena que los responsables de la toma de decisiones asuman ese riesgo legal para reducir el sesgo total?

Incluso si un algoritmo aborda las preocupaciones sobre la posible discriminación en un proceso de RRHH, existen además preocupaciones sobre la justicia individual. Supongamos que un algoritmo determinó los dos mejores candidatos con un 80% y 90% de coincidencia con el perfil de mejor rendimiento. ¿La diferencia del 10% entre ambos es grande o pequeña, teniendo en cuenta la precisión del algoritmo? El responsable de selección podría revisar toda la información disponible y emitir un juicio experto. Sin embargo, hay suficientes pruebas de que tales juicios incluyen a menudo racionalizaciones e introducen también los propios sesgos del individuo, cuando no están afectados por la arbitrariedad.

Una manera de mitigar algunos de estos problemas es introducir la variación aleatoria. La investigación muestra que los empleados perciben un proceso aleatorio como justo para determinar decisiones complejos y por lo tanto inciertas.(Lind y Van den Bos 2002). Si el algoritmo es sólido y ambos candidatos son fuertes, tiene más sentido aceptar la complejidad de las decisiones de contratación y hacer una elección aleatoria entre los dos candidatos, posiblemente, con las probabilidades de selección proporcional a las puntuaciones correspondientes. En otras palabras, la aleatorización puede ser una herramienta de gestión de la IA tanto en la etapa de toma de decisiones como en las etapas anteriores del ciclo de vida de la IA.

Como siempre, el contexto importa. En general, los responsables de RRHH encuentran aceptable el uso de algoritmos para tomar decisiones que esencialmente recompensan a los empleados (a quién promover, a quién contratar en primer lugar etc…) Sin embargo, el uso de algoritmos predictivos para "castigar" a los empleados plantea más interrogantes. Un algoritmo que predice los rendimientos futuros de los empleados seguramente se introducirá en algún momento para tomar decisiones de despido, pero ¿qué hay de uno que prediga quién robará a la compañía o cometerá un delito?

Cuestiones éticas.

Aquí nos enfrentamos a un dilema ético. La noción utilitarista se pondrá siempre a favor del uso de algoritmos predictivos para eliminar problemas y empleados que estén menoscabando el bien del común. Pero una posición quizás más kantiana, mantendrá que los individuos deben ser juzgados en base a sus "acciones" y encontraría esa posición altamente censurable.

Cuestiones de compromiso de gestores y empleados.

Relacionado con el concepto de justicia nos encontramos con el problema de la aceptabilidad de las decisiones tomadas por algoritmos y de la participación e involucración de las personas en dichas decisiones.

La investigación muestra cada vez más que los algoritmos funcionan mejor que el juicio humano cuando se utilizan para predecir resultados repetitivos, como la lectura de los rayos X, pero también para predecir los resultados sobre los empleados o los candidatos a un puesto de trabajo (Cowgill 2018). Pero si los algoritmos se hacen cargo de la contratación y los supervisores no juegan ningún papel en el proceso, ¿estarán tan

comprometidos con las nuevas contrataciones como si hubieran tomado ellos mismos las decisiones de contratación?

El reto de la "explicabilidad" del modelo de aprendizaje automático (Machine Learning)

Con los modelos de aprendizaje automático, la capacidad de explicar qué factores está utilizando el algoritmo para hacer predicciones es generalmente mucho más difícil que con los modelos estadísticos tradicionales. Una simple regla de decisión sobre la antigüedad - los trabajadores de mayor antigüedad tienen preferencia sobre los de menor antigüedad - es fácil de entender y se percibe como objetiva aunque no nos gusten sus implicaciones.

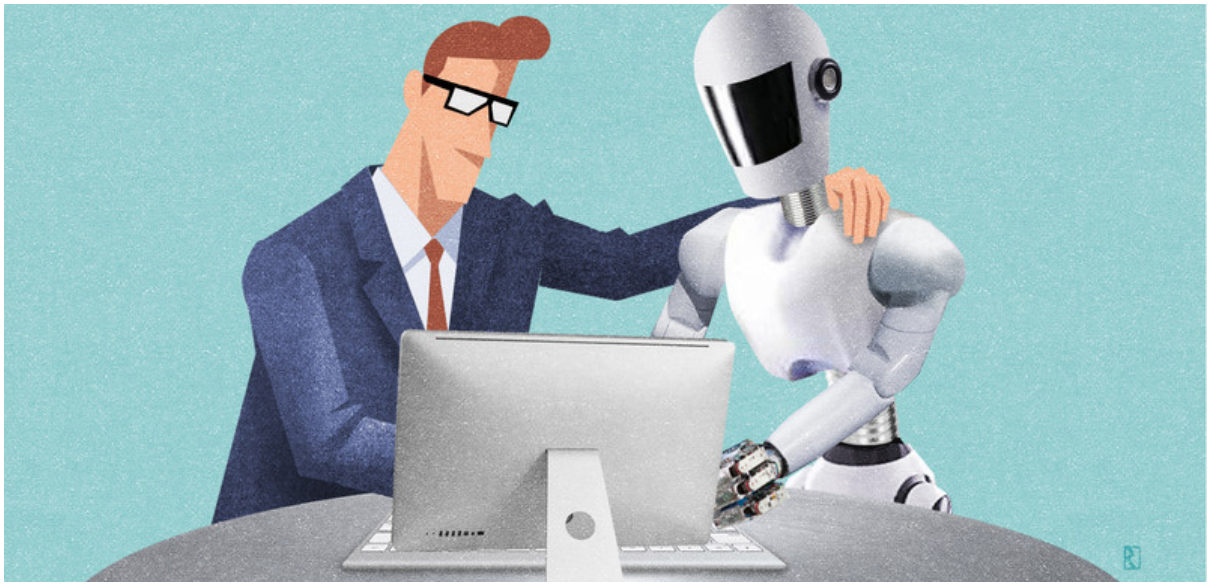
Un algoritmo de ML basado en una combinación ponderada de 10 factores relacionados con el rendimiento puede ser tan objetivo como un modelo estadístico tradicional, pero es mucho más difícil de entender, especialmente cuando los empleados hacen entre ellos comparaciones inevitables. Los algoritmos son más precisos cuanto más complejos son, pero también más difíciles de entender y explicar.

Un ejemplo bien conocido de la importancia de la explicabilidad para los usuarios proviene de la aplicación del programa para oncología de IBM Watson. Esta aplicación encontró una resistencia considerable por parte de los oncólogos porque era difícil entender cómo el sistema estaba llegando a sus decisiones. Cuando el juicio emitido por el programa no estaba de acuerdo con la evaluación del médico, esta falta de transparencia dificultó que los expertos médicos aceptaran y actuaran de acuerdo con las recomendaciones que el sistema producía.

Especialmente en contextos en los que hay "mucho en juego", como los que afectan a la vida de las personas o a sus carreras, es probable que la explicabilidad se convierta en un imperativo para el uso exitoso de las tecnologías de aprendizaje automático.

Reacciones de los empleados a las decisiones algorítmicas.

Los procesos de toma de decisiones formales afectan inevitablemente al comportamiento de los empleados. En este sentido, podemos aprender mucho de los esfuerzos de la "Dirección Científica" para desarrollar reglas de decisión óptimas. Las prácticas de empleo y las decisiones sobre la organización del trabajo se basaban en principios de ingeniería y experimentos humanos a priori. Aunque las decisiones así tomadas pueden haber sido mucho más eficientes que las prácticas anteriores, generaban un amargo resentimiento en los trabajadores, lo que inevitablemente desembocaba en conflictos entre los trabajadores y la dirección.



Desde el punto de vista de los empleados de primera línea y sus supervisores, la utilización del modelo de IA por parte de la dirección de la empresa para la toma de decisiones de RRHH no es muy distinta. Las decisiones se percibirían como ajenas a ellos mismos, tomadas por una máquina a la que se le presupone una gran capacidad de predicción y hasta cierto punto una infalibilidad. Tratar de oponerse a sus decisiones es ir en contra de la ciencia y la tecnología.

Para ilustrar una consecuencia probable del uso de la IA en RRHH, tenemos que contemplar la relación que se establece entre empleado y supervisor como el auténtico generador del rendimiento. La relación con el supervisor es crucial para el desempeño de sus subordinados y la calidad de esa relación depende del intercambio social: "Yo como supervisor te cuido, y tú como subordinado haces bien tu trabajo." Incluso cuando los empleados tienen poco compromiso con su empleador como organización, pueden sentir compromiso con su supervisor. ¿Cómo se ve afectado este intercambio cuando las

decisiones que antes tomaba mi supervisor son tomadas ahora por un algoritmo o incluso en gran medida por un algoritmo en lugar de por un supervisor?

En un contexto laboral, si mi supervisor me asigna a trabajar otro fin de semana de este mes, algo que no quiero hacer, podría hacerlo sin quejarme si creo que mi supervisor ha sido justo conmigo. Incluso podría sentir empatía con el aprieto en el que se encuentra mi supervisor cuando tiene que llenar el turno del fin de semana. Sin embargo, cuando mi horario de trabajo es generado por el software, no tengo ningún tipo de "buena voluntad" hacia ese programa, y no puedo sentir empatía con él. Tampoco puedo quejarme a la máquina. Sabemos, por ejemplo, que las personas responden de manera muy diferente a las decisiones tomadas por algoritmos que a las decisiones tomadas por personas (Dietvorst, Simmons y Massey 2016). Si hay buenas noticias, como una bonificación, se establece una relación con mi supervisor si parece que al menos ha estado involucrada en la decisión, algo que no sucede si esa decisión es generada por un algoritmo.

Sin embargo, puede haber ocasiones en las que las decisiones son más fáciles de aceptar cuando son tomadas por un algoritmo que cuando son tomadas por un humano, especialmente cuando esas decisiones tienen consecuencias negativas para nosotros. Los usuarios de Uber, por ejemplo, responden negativamente a los aumentos de precios cuando perciben que están establecidos por un ser humano (tratando de explotarlos) en lugar de por un algoritmo.

En resumen.

Mientras que la IA de propósito general es todavía una posibilidad remota en cualquier campo de la actividad humana, la velocidad de progreso hacia los sistemas especializados de IA en el cuidado de la salud, la industria automotriz, las redes sociales, la publicidad y el marketing es considerable. Por contra, se ha avanzado mucho menos en cuestiones relacionadas con la gestión de los empleados. Identificamos cuatro razones: la complejidad de los fenómenos de RRHH, los pequeños datos de las operaciones de RRHH, la imparcialidad y las restricciones legales y las reacciones de los empleados a la gestión de la IA.

El "razonamiento causal" es el primer principio relevante para abordar estos desafíos en todas las etapas del Ciclo de Vida de la IA, ya que nos permite extraer conclusiones prácticas de datos pequeños y facilita la toma de decisiones basadas en algoritmos de forma más justa y explicable.

Pero estos beneficios no vienen sin costes: las empresas tienen que aceptar el menor poder predictivo de estos algoritmos y buscar algún consenso sobre las suposiciones causales que se introducen en ellos.

La “aleatorización” es un segundo principio que puede ayudar a establecer la causalidad y también puede ser una regla de decisión para cuando nos enfrentamos a un bajo poder predictivo de los algoritmos.

Por último, es necesario formalizar los procesos para garantizar que los supuestos incorporados en los algoritmos sean razonables y que los resultados sean comprensibles.

En qué medida todos estos cambios requieren una reestructuración de la función de RRHH es una cuestión importante. Ciertamente, los líderes de RRHH necesitan entender y facilitar las etapas de implantación de la IA en su área. Por otro lado, la integración de los datos de RRHH con los datos de negocio y financieros debería permitir a un departamento de RRHH cuantificar en términos monetarios su contribución a los resultados de la empresa.

Los supervisores y gerentes de línea también tendrán que actualizar sus habilidades. Para ellos, la IA más que inteligencia artificial debe significar "inteligencia aumentada” y suponer un uso informado de la tecnología aplicada a la toma de decisiones.

Bibliografía de interés.

Barends, Eric and Denise M. Rousseau. 2018. Evidence-Based Management: How to Use Evidence to Make Better Organizational Decisions. Kogan Page.

Bock, Laslo. 2015. Work Rules! Insights from Inside Google That Will Transform How You Live and Lead. Hachette Book Group.

Cappelli, Peter. 2017. “There’s No Such Thing as Big Data in HR.” Harvard Business Review. June.

Cappelli, Peter and AnnaTavis. 2017. The Performance Management Revolution. Harvard Business Review, November.

Denrell, Jerker, Christina Fang, Chengwei Liu. 2015. “Change Explanations in Management Science.” Organization Science 26(3): 923-940.

IBM 2018.Cowgill, Bo (2017) The Labor Market Effects of Hiring through Machine Learning “Bias and Productivity in Humans and Algorithms.

Cowhill, “Theory and Evidence from Résumé Screening.” Working paper.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2016). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. Management Science, 64(3), 1155-1170.

Jennifer M. Logg, Julia A. Minson, Don A. Moore. Algorithm Appreciation: People Prefer Algorithmic To Human Judgment. Working Paper 17-086 HBS.

Junqué de Fortuny, E., Martens, D., & Provost, F. (2013). "Predictive modeling with big data: is bigger really better?" Big Data, 1(4), 215-226.

Lee, M. K., Kusbit, D., Metsky, E., & Dabbish, L. 2015. Working with machines: The impact of algorithmic, data-driven management on human workers. Proceedings of the 33rd Annual ACM SIGCHI Conference: 1603-1612. Begole, B., Kim, J., Inkpen, K & Wood, W (Eds.), New York, NY: ACM Press.

Lind, E. Allan and Kees Van den Bos. 2002. "When Fairness Works: Toward a General Theory of Uncertainty Management." Research in Organizational Behavior 24: 181-223.

LinkedIn. 2018. The Rise of HR Analytics.

Loftus, Joshua R., Chris Russel, Matt J. Kusner, and Ricardo Silva. "Causal Reasoning for Algorithmic Fareness." arXiv:1805.05859

Malinsky, Daniel and David Danks. 2017. "Causal Discovery Algorithms: a Practical Guide." Philosophy Compass https://doi.org/10.1111/phc3.12470.

Pearl, Judea. 2018. The Book of Why: The New Science of Cause and Effect. Basic Books.

Pfeffer, Jeffrey and Robert I. Sutton. 2006. Hard Facts, Dangerous Half-Truths and Total Nonsesne: Profiting from Evidence-Based Management. Harvard Business Review Press.

Rousseau, Denise (Editor). 2014. The Oxford Handbook of Evidence-Based Management. Oxford University Press.

Tucker, Catherine. 2017. "Privacy, Algorithms, and Artificial Intelligence." In The Economics of Artificial Intelligence: An Agenda. Edited by Ajay K. Agrawal,

Sobre el autor:

José María Torán es desde 2008 socio de CCG, donde ha desarrollado proyectos de organización comercial, diseño de procesos comerciales y mejora del rendimiento humano en sectores como gran consumo, retail y sector financiero. Durante su periodo en Clarity se ha interesado especialmente por desarrollar actividades que permitan trasladar la estrategia de las compañías a los comportamientos diarios de los integrantes de la organización.

Sobre Clarity Consulting Group.

Somos una firma de servicios profesionales dedicada a mejorar la eficacia de las organizaciones y la rentabilidad de nuestros clientes.

Estamos especializados en conquistar el territorio entre la definición de la estrategia y la consecución efectiva de resultados, alineado la organización hacia lo que realmente importa: satisfacer las necesidades de los clientes de manera diferencial, mejorando las capacidades de la organización para conseguir resultados.

Clarity Consulting Group

Felix Boix 18 - 1º izda. Madrid 28036.

www.claritygroup.es