

Medicina personalizada y ciencia de datos.

El uso de analíticas de datos para la selección de tratamientos

Gustavo **Mendoza**
Josefa **Somodevilla**
Concepción **Pérez de Celis**

En los últimos años, el uso de las tecnologías de la información y comunicación ha incrementado los datos que se producen y almacenan a cada momento. La Inteligencia de Negocios (*Business Intelligence*) ha usado desde su origen información histórica para la toma de decisiones en las empresas, presentando respuestas a las preguntas ¿qué pasó?, ¿por qué pasó?, ¿qué pasará? y ¿qué decisión debe tomarse? Actualmente, este proceso se ha extrapolado hacia otras ramas de la ciencia, demostrando su relevancia y los beneficios que ofrece para la generación del conocimiento. ¿En qué consiste este proceso conocido como “ciencia de datos” y cómo se ha comenzado a aplicar en áreas como el cuidado de la salud, retomando la idea del paradigma de la medicina personalizada?

EL PARADIGMA DE LA MEDICINA PERSONALIZADA

Los términos Medicina Personalizada o Medicina de Precisión comúnmente son utilizados como sinónimos. Para la FDA (Food and Drug Administration), esta visión de la



Figura 1. Información necesaria para una real implementación de la medicina personalizada. Fuente: Goetz y Schork, 2018.

medicina se basa en la premisa de proveer “al paciente adecuado: el medicamento correcto, en la dosis correcta, en el tiempo correcto” (Pirracchio *et al.*, 2019). Con los recientes descubrimientos acerca del genoma humano y la farmacogenética (una rama de la genética que estudia cómo la efectividad de los medicamentos es afectada por la variabilidad genética), la capacidad de almacenamiento de información en la nube y la capacidad de las computadoras actuales para procesar información, la idea de la medicina personalizada se vislumbra como el siguiente paso a dar en el cuidado de la salud.

Aunque pueda parecer una idea relativamente moderna, en realidad este paradigma fue propuesto hace más de un siglo por Archibald Garrod, un médico inglés que estudiaba algunas enfermedades raras como alcaptonuria, albinismo, cetonuria y pentosuria. Tras analizar sus resultados, Garrod concluyó que, al menos en el metabolismo, los seres humanos tenían una gran variabilidad y que esas diferencias en el metabolismo podrían ayudar a explicar evidentes diferencias entre individuos, así como en la susceptibilidad de los individuos a algunas enfermedades y en la manera como se

manifiestan (Goetz y Schork, 2018). De acuerdo con este paradigma, la información que debe ser recolectada y procesada para una implementación real de medicina personalizada, se presenta en la Figura 1.

El acceso al cuidado de la salud es importante, pues puede ser que el paciente no tenga acceso a especialistas o tecnología debido a razones geográficas o económicas y, por tanto, las intervenciones deben ser elaboradas tomando en cuenta esta limitación. La información genética hereditaria es útil en situaciones predictivas o de diagnóstico; sin embargo, cambios somáticos en el ADN pueden proveer información importante en procesos patogénicos.

Los biomarcadores y muestras son útiles para detectar cambios en la salud, aunados a la imagenología, la radiología y los datos que puedan recolectarse mediante dispositivos inalámbricos. La exposición ambiental, los hábitos y la personalidad realmente pueden impactar en los resultados de una intervención. Finalmente, las modificaciones epigenéticas (aquellas que cambian el funcionamiento de los genes sin cambiar la secuencia genética) también deben ser monitoreadas, pues pueden indicar un cambio en el estado de la salud.

Es evidente que los requerimientos de información son exigentes y su sola recolección significa ya un reto. Adicionalmente, representa un problema para los médicos el analizar y procesar en un tiempo eficiente toda esta información para un solo paciente. Es aquí donde el proceso de ciencia de datos puede utilizarse en busca de patrones y relaciones entre los datos que permitan a los profesionales de la salud tomar decisiones basadas en el contexto personal del paciente, reduciendo el riesgo de sesgos personales y experienciales.

Actualmente, existen algunos acercamientos hacia este paradigma dentro de la práctica médica; por ejemplo, algunos medicamentos como la warfarina (que se utiliza como anticoagulante) y el imatinib

(empleado en algunos tipos de cáncer) aparentemente solo funcionan (o no presentan efectos secundarios) en pacientes que poseen ciertos perfiles genéticos, lo que ha motivado el interés en identificar los factores genéticos que influyen en las respuestas individuales de los pacientes a diferentes fármacos e intervenciones. Estos acercamientos se han realizado en tres líneas principales de investigación:

- Terapias de mutación específica: el medicamento llamado ivacaftor es utilizado en casos de fibrosis quística en personas que tienen una mutación específica en el gen CTRF (Coalition, 2018). Otro ejemplo es el conjunto de tratamientos para cáncer llamado “inmunoterapias”. Este tipo de tratamientos apunta a utilizar el sistema inmunológico propio del paciente para atacar al cáncer mediante la utilización de células T modificadas para reconocer y atacar tumores con características genómicas específicas, llamadas “neoantígenos”. Aunque han demostrado tener buenos resultados, son tratamientos altamente personalizados pues, si el tumor no tiene la firma de neoantígenos correcta, entonces las células T no podrán afectarlo. Adicionalmente, no pueden utilizarse células T de otra persona, pues pueden tener un comportamiento agresivo (Farkona *et al.*, 2016).
- Estrategias de detección temprana personalizada: al tener disponible la información genómica de una persona, es posible identificar si tiene susceptibilidad de padecer alguna enfermedad, lo que permite establecer márgenes de referencia personalizados de indicadores específicos, en vez de los indicadores generales de la población.
- Prevención personalizada de enfermedades: el uso de la información genética puede ayudar también a prevenir ciertas enfermedades. Por ejemplo, el trabajo de Liao y su equipo reportó una mejora

en la supervivencia y un decremento en el riesgo de muerte de pacientes con cáncer colorrectal que tienen una mutación somática en el gen PIK3CA (Liao *et al.*, 2012). Sobre esta misma enfermedad, Nan y su equipo reportaron que el uso de aspirina puede reducir la posibilidad de desarrollar este tipo de cáncer en personas con un determinado genotipo (Nan *et al.*, 2015).

¿Cuándo debe personalizarse un tratamiento para un paciente?, la respuesta a esta pregunta no es tan simple como parece. Si se tiene acceso a la información necesaria del paciente, al conocimiento y a la capacidad de cómputo requerida para tomar las decisiones en un tiempo aceptable para el profesional de salud y el paciente, entonces la medicina personalizada debería ser el paradigma para utilizar en la práctica médica. Sin embargo, aún hay brechas importantes por cubrir.

EL PROCESO DE LA CIENCIA DE DATOS

Se le llama ciencia de datos al enfoque multidisciplinario utilizado para obtener indicadores (conocidos en inglés como *insights*) a partir de conjuntos de datos creados, almacenados y que crecen día con día (Poornima y Pushpalatha, 2020).

Este proceso está compuesto de cinco fases principales reconocidas, como se muestra en la Figura 2.

Si bien cada fase es importante y relevante dentro del proceso de ciencia de datos, para efectos de la obtención de indicadores se hará referencia solo a la fase del procesamiento de datos. Es en esta fase donde se aplican técnicas de *machine learning* (ML) sobre conjuntos de información para obtener indicadores significativos y relevantes acerca de los mismos. Se le llama analítica al conjunto de indicadores resultantes tras aplicar técnicas de ML con el

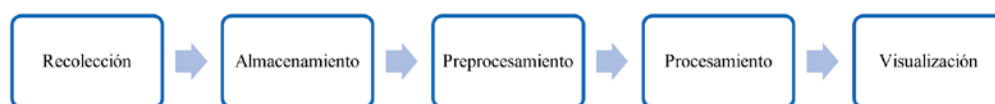


Figura 2. El proceso de ciencia de datos.

objetivo de analizar la información en un nivel específico de profundidad, y se clasifica de acuerdo con el grado de conocimiento que genera, como se muestra en la Figura 3.

En el nivel más bajo de analíticas se encuentran las descriptivas; estas trabajan sobre grandes cantidades de datos históricos para dar respuesta a la pregunta ¿qué pasó? y, en un grado de sofisticación más elevado, describen la causa del fenómeno si es que se llega al proceso de diagnóstico (¿por qué pasó?). Sus procesos son tomados principalmente de la estadística descriptiva y permiten obtener analíticas que servirán de base para el análisis predictivo.

En el siguiente nivel de analíticas, las predictivas toman el resultado de las analíticas descriptivas para construir un modelo que permita realizar predicciones que respondan a la pregunta ¿qué pasará? Los procesos utilizados en este nivel de análisis son variados, pues hay desde aplicaciones de la estadística inferencial, hasta procesos multidisciplinarios como las redes neuronales o la minería de datos. En el nivel más alto de analíticas se encuentran las prescriptivas, que toman los resultados de los procesos anteriores para enfocarse en la toma de decisiones, respondiendo a la pregunta ¿qué es lo que se debe hacer? Para esto, se basan en técnicas como la optimización, la simulación o la ciencia de redes para evaluar todos los posibles escenarios que el modelo predictivo permite construir, y justifican sus resultados considerando la incertidumbre.

Es en este nivel de análisis, el prescriptivo, donde la medicina personalizada se perfila como un campo de aplicación. Los datos históricos sobre los que deberá trabajar el proceso de ciencia de datos serán aquellos obtenidos de expedientes clínicos de las propias instituciones de salud. En una visión limitada, cada institución puede trabajar con sus

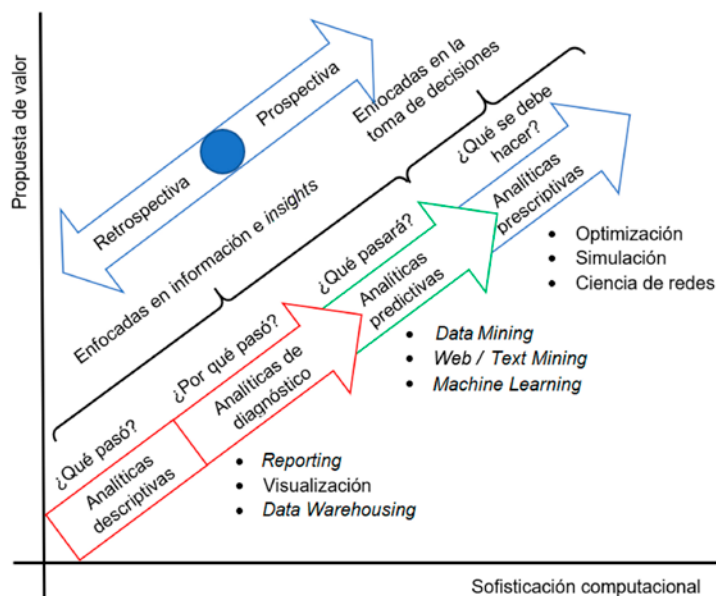


Figura 3. Proceso de obtención de analíticas. Fuente: Mosavi y Santos, 2020.

propios datos o, como sucede en algunos países actualmente, trabajar con los datos de sus sistemas centralizados de salud para tener una visión multicéntrica de la problemática a analizar. Por medio de las analíticas descriptivas se pueden obtener, por ejemplo, clasificaciones de tratamientos previamente prescritos por los profesionales de salud para un determinado padecimiento, de acuerdo con determinadas características del paciente, de manera que pueda conformarse una muestra representativa a partir de la cual se construya un modelo matemático que permita inferir cómo se relacionan esas características con el tratamiento previamente prescrito por los profesionales.

Este modelo se evalúa con un conjunto de datos diferente al utilizado para su construcción (pero, también, evaluados y calificados por profesionales), de manera que pueda medirse la precisión al comparar los tratamientos obtenidos por el modelo contra los prescritos por los médicos que evaluaron y calificaron los datos de prueba.

Finalmente, este modelo será dotado de valores de probabilidad obtenidos directamente de la literatura médica para sugerir el resultado que garantice un valor óptimo, de acuerdo con el criterio que se busque optimizar.

EJEMPLO DE APLICACIÓN: CÁNCER DE MAMA

Una de las principales enfermedades crónicas a nivel mundial es el cáncer. De acuerdo con datos de la Organización Mundial de la Salud (OMS), esta enfermedad fue la principal causa de muerte de 2020, con 10 millones de defunciones, de las cuales 2.26 millones fueron nuevos casos de cáncer de mama (WHO, 2022).

El término “cáncer” engloba a un amplio grupo de enfermedades que pueden afectar y presentarse en cualquier parte del organismo. Su principal característica es la multiplicación rápida de células anormales que se extienden más allá de sus límites habituales y pueden invadir partes adyacentes del cuerpo o extenderse a otros órganos en un proceso llamado metástasis. Es entendible el interés en la investigación de nuevos métodos de detección temprana y tratamiento.

En México, la principal causa de muerte por cáncer en mujeres es el cáncer de mama; esto representa, de acuerdo con el Instituto Nacional de Salud Pública, la muerte de una mujer cada dos horas en 2022, a pesar de ser un cáncer que es curable si se trata en forma temprana. De acuerdo con el gobierno mexicano, los principales motivos de que este tipo de cáncer se vuelva mortal son un diagnóstico tardío por parte de los médicos y un inicio tardío del tratamiento por parte de los pacientes. Los protocolos de tratamiento para este padecimiento se encuentran recopilados y estandarizados en el *Consenso mexicano sobre diagnóstico y tratamiento del cáncer mamario*, el cual es un documento elaborado por la Sociedad Mexicana de Oncología A. C. revisado y actualizado anualmente, que sirve como herramienta auxiliar en la práctica diaria de los oncólogos y como fundamento de la Norma Oficial Mexicana para el tratamiento de este padecimiento (Cárdenas-Sánchez *et al.*, 2021).

Dos ejes deben seguirse para la creación de tratamientos personalizados para cáncer de mama: uno es centrado en el paciente y otro centrado en la información existente, como se muestra en la Figura 4.

En el eje centrado en los datos se trabaja con estudios retrospectivos basados en información

tomada de expedientes clínicos previamente anonimizados, y siguiendo los requerimientos de seguridad que las leyes y los propietarios de los expedientes determinen.

Sobre estos datos, tomando como referencia la información presente en el consenso, el primer paso consiste en obtener analíticas descriptivas que permitan conocer qué tratamientos han dado los médicos para cada subclasificación de cáncer de mama, agrupándolos y categorizándolos de manera que puedan construirse reglas que deberán coincidir con las especificadas por el consenso, y evaluar parámetros como su resultado (remisión o deceso), tiempo de remisión, recurrencia, etcétera.

Una vez obtenidas las analíticas descriptivas, el análisis predictivo deberá ser aplicado mediante técnicas que permitan entender la decisión tomada como resultado, y qué indicadores están siendo tomados en cuenta y cuáles no. Este modelo deberá ser avalado siempre por un médico.

En el eje centrado en el paciente, este debe ser copartícipe en la recolección de la información adicional necesaria para la construcción de su contexto. Esto puede ser hecho mediante herramientas como registros personales de salud que faciliten la comunicación entre paciente y médico; o bien, ser recopilada por los profesionales de salud en el ejercicio de su intervención.

Figura 4. Ejes para creación de tratamientos personalizados.



La personalización del tratamiento del paciente será realizada en el análisis prescriptivo, el cual puede ser llevado a cabo mediante técnicas como el aprendizaje por reforzamiento, que es una técnica basada en el modelo de aprendizaje de prueba y error. En esta técnica, una entidad independiente capaz de tomar decisiones (que recibe el nombre de agente) es dotada de un conjunto de reglas básicas con las cuales debe modificar su entorno, de manera que llegue a un estado final. Cada cambio al entorno otorga una recompensa o penalización al agente, que va agregando nuevas reglas de acuerdo con las interacciones con el entorno.

El objetivo es llegar al estado final optimizando el valor de la recompensa. El entorno será el contexto personal y clínico del paciente, y el agente se encargará de evaluar la probabilidad de éxito del tratamiento a lo largo del tiempo.

CONCLUSIONES

Es importante hacer notar que no se propone sustituir la labor médica por herramientas computacionales, sino otorgar herramientas basadas en datos y en literatura médica para que, tanto los profesionales, como los pacientes, se vean beneficiados. Por el lado de los profesionales, en el desarrollo de herramientas que faciliten la toma de decisiones de tratamientos que ofrezcan mayores probabilidades de efectividad, de acuerdo con el contexto del paciente; por el lado del paciente, otorgando herramientas enfocadas hacia la detección temprana, el seguimiento, la adherencia al tratamiento y recomendaciones basadas en evidencias científicas para el aspecto paliativo y emocional de la enfermedad.

Con una visión multidisciplinaria de la problemática, es posible que los tratamientos personalizados puedan traducirse en mejores posibilidades de recuperación y un mayor nivel de calidad de vida para los pacientes, con los consecuentes beneficios que una población sana representa para un país.

REFERENCIAS

- Cárdenas-Sánchez J, Bargalló-Rocha E, Erazo Valle A, Poitevin Chacón A, Valero Castillo V y Pérez Sánchez V (2021). Consenso Mexicano sobre diagnóstico y tratamiento del cáncer mamario. *Gaceta Mexicana de Oncología* 2-55. <https://www.elsevier.es/es-revista-gaceta-mexicana-oncologia-305-articulo-consenso-mexicano-sobre-diagnostico-tratamiento-X1665920115428989>.
- Coalition PM (2018). *Personalized Medicine at FDA: 2017 Progress Report*.
- Farkona S, Diamandis EP and Blasutig IM (2016). Cancer immunotherapy: The beginning of the end of cancer? *BMC Medicine* 14:73. <https://doi.org/10.1186/s12916-016-0623-5>.
- Goetz LH and Schork NJ (2018). Personalized medicine: Motivation, challenges, and progress. *Fertility and Sterility* 109(6):952-963. <https://doi.org/10.1016/j.fertnstert.2018.05.006>.
- Liao X, Lochhead P, Nishihara R *et al* (2012). Aspirin use, tumor PI-K3CA mutation, and colorectal-cancer survival. *New England Journal of Medicine* 367(17):1596-1606.
- Mosavi N and Santos M (2020). How Prescriptive Analytics Influences Decision Making in Precision Medicine. *Procedia Computer Science* 177:528-533. <https://doi.org/10.1016/j.procs.2020.10.073>.
- Nan H, Hutter CM, Lin Y *et al* (2015). Association of aspirin and NSAID use with risk of colorectal cancer according to genetic variants. *Jama* 313(11):1133-1142.
- Pirracchio R, Cohen MJ, Malenica I, Cohen J, Chambaz A, Cannesson M, Lee C, Resche-Rigon M and Hubbard A (2019). Big data and targeted machine learning in action to assist medical decision in the ICU. *Anaesthesia Critical Care & Pain Medicine* 38(4):377-384. <https://doi.org/10.1016/j.accpm.2018.09.008>.
- Poornima S and Pushpalatha M (2020). A survey on various applications of prescriptive analytics. *International Journal of Intelligent Networks* 1:76-84. <https://doi.org/10.1016/j.ijin.2020.07.001>.
- WHO (2022). *Cáncer*. <https://www.who.int/es/news-room/fact-sheets/detail/cancer>.

**Facultad de Ciencias de la Computación
Benemérita Universidad Autónoma de Puebla**

Gustavo Mendoza
gustavo.mendozao@alumno.buap.mx

Josefa Somodevilla
maria.somodevilla@correo.buap.mx

Concepción Pérez de Celis
maria.perezdecelis@correo.buap.mx