



# Big Data: el dato al servicio del negocio



# Big Data: el dato al servicio del negocio

Big Data ayuda a procesar, analizar y obtener el máximo valor de todos los grandes volúmenes de información existentes en múltiples formatos y provenientes de distintas, permitiendo tomar decisiones inteligentes de manera rápida y rentable. Conscientes de ello, empresas como GMV están siendo tractoras en la utilización de estas tecnologías, desarrollando ambiciosos proyectos en ámbitos diversos, como el bancario, energético o sanitario.

Según el informe Enterprise Data Driven, Big Data es la segunda prioridad para las empresas en cuanto a inversión en nuevas tecnologías en los próximos 12 meses, sólo por detrás de la movilidad, pero por delante de iniciativas como Cloud e Internet de las Cosas.

IDC estima un crecimiento anual compuesto del 11,7 por ciento en la inversión en tecnologías relacionadas con la gestión y análisis de los datos, pasando de los 130.100 millones de dólares que se dedicaron a esta tarea en 2016 a más de 203.000 millones para el final de la década.

Para poder analizar lo que supone Big Data y cómo sacar provecho, hemos querido conversar con José Carlos Baquero, jefe de División de Ingeniería de Software y Desarrollo en GMV Secure e-Solutions, quien nos explica que el de Big Data “es un mercado maduro, empieza a ser una tecnología commodity. La mayoría de las empresas conocen lo que les puede ofrecer Big Data, pero queda recorrido para extraer mayor valor de

los datos, ya que, básicamente, los datos se están recogiendo y almacenando en un data lake. Es el momento para realizar predicciones, optimizaciones y recomendaciones a partir de ellos”, y añade que, gracias a los casos que se han hecho públicos recientemente, “empieza a haber una mayor concienciación sobre la seguridad y privacidad de los datos”.

Tradicionalmente juntas en listas de tendencias tecnológicas, Big Data y Cloud son tecnologías tendentes a converger porque, como señala nuestro interlocutor, Cloud permite realizar pruebas de concepto de forma rápida y con costes contenidos, además de aportar una capacidad de proceso elástica, sobre todo a la hora de

***Según el informe Enterprise Data Driven, Big Data es la segunda prioridad para las empresas en cuanto a inversión en nuevas tecnologías en los próximos 12 meses, sólo por detrás de la movilidad***





## Big Data: el dato al servicio del negocio



mejora de procesos internos, motores de recomendación...), si bien en otros como Sanidad (análisis de imágenes, diagnóstico, monitorización, estudios clínicos, genética...) las posibilidades son enormes; al igual que en Automoción (coches autónomos), Industria 4.0 (optimización de procesos industriales, eficiencia energética...) o Ciberseguridad”.

Por último, hablando de retos que debe asumir la tecnología Big Data, nos indica algunos tales como “Edge

Computing, Deep Learning con capacidades cognitivas, gestión de datos personales por los usuarios, seguridad y privacidad (Security by Default and Security by Design; Privacy by Default and Privacy by Design), integración con la IA para ayudar en las tareas más repetitivas y de poco valor, como la transcripción de documentos o clasificación de archivos, permitiéndonos centrarnos en tareas con mayor grado intelectual e imaginación”.

### *IDC estima un crecimiento anual compuesto del 11,7 por ciento en la inversión en tecnologías relacionadas con la gestión y análisis de los datos*

entrenar los algoritmos de Inteligencia Artificial. Asimismo, ofrece almacenamiento de bajo coste, alineamiento de los costes de la infraestructura con el valor que ofrece la plataforma e innovación fácilmente aprovechable”, pese a que surgen reticencias por motivos de seguridad y privacidad. Recordemos que los principales riesgos percibidos en Big Data son la confidencialidad, la privacidad, las auditorías y el cumplimiento normativo.

Hablando de sectores, y en estas páginas vemos ejemplos de muchos de ellos, Big Data está ofreciendo grandes beneficios “en aquellos digitalizados desde hace ya más tiempo, como Banca, Seguros, Teleco... (fraude,



José Carlos Baquero, líder de las iniciativas de Big Data en GMV



**PINCHA SOBRE LA IMAGEN PARA VER LA ENTREVISTA**

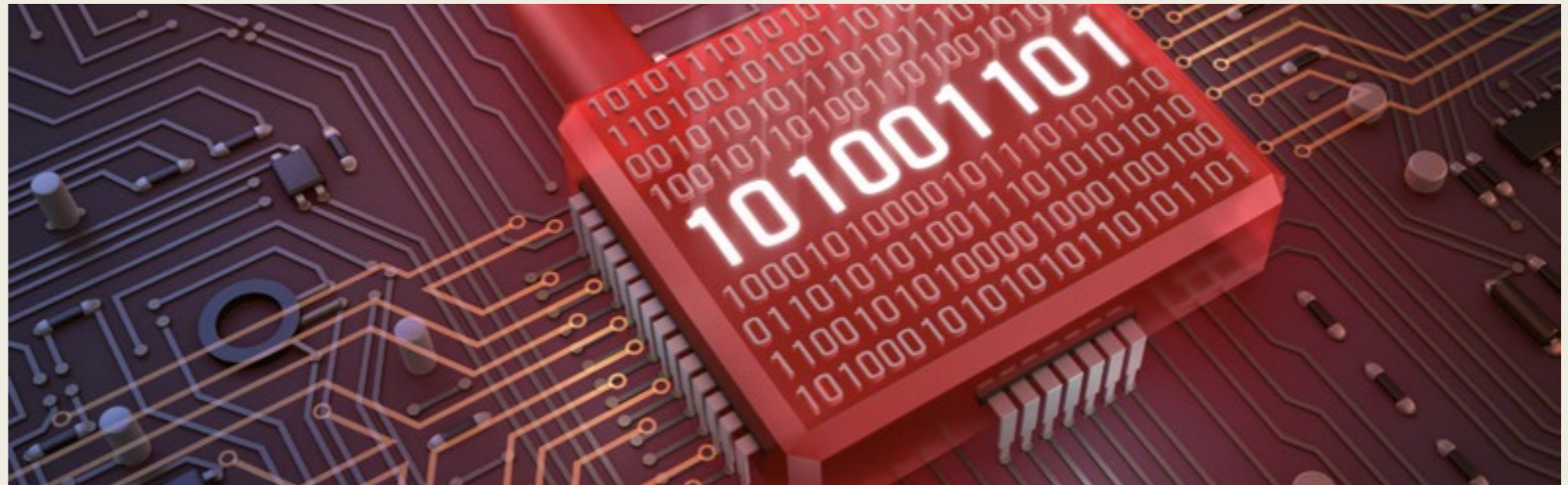
## El uso de la Inteligencia Artificial en la prevención del fraude

La creciente digitalización de los negocios conlleva una mayor exposición al fraude.

- El acceso a productos y servicios en línea añaden vectores de ataque a los ya tradicionales. El fraude interno (cometido por empleados y proveedores, conscientemente o no) tiene un impacto cada vez mayor.
- Los nuevos servicios y modelos de negocio son aprovechados por los delincuentes para crear nuevas formas de fraude.
- La inmediatez de respuesta que imponen los clientes, así como la dificultad de rastrear al autor de un fraude, suponen un hándicap adicional. Tenemos segundos, minutos, unas pocas horas, para frenar un posible fraude.

Tradicionalmente el fraude se ha analizado “manualmente”, revisando los casos mediante reglas preestablecidas. Un proceso laborioso, caro, lento, ineficiente y sujeto a errores. ¿Por qué dejar a los humanos lo que puede hacer de manera mucho más eficiente una máquina?

*La prevención del fraude ha encontrado en el Big Data y la Inteligencia Artificial unos aliados de indudable valor*



La prevención del fraude ha encontrado en el Big Data y la Inteligencia Artificial unos aliados de indudable valor. La combinación de ambos permite cruzar ingentes volúmenes de información (en tiempo real o no), descubrir patrones y relaciones y conexiones ocultas en ella, partiendo de fuentes muy diversas.

Por poner ejemplos:

- En el sector bancario, es posible la detección de operaciones fraudulentas con tarjeta, cuentas mulla o solicitudes de crédito con riesgo de impago, entre otras muchas.
- En el sector de los seguros, se puede detectar fraude por simulación de partes o connivencia con terceros (talleres, clínicas...), así como ventas abusivas o modificación de pólizas en beneficio propio o de familiares, por parte de empleados.
- En el ecommerce, el fraude en contracargos (chargeback).

Ya no sólo se trata de responder ante los casos de fraude, sino de ser proactivos y ágiles en la respuesta, protegiendo contra el fraude y detectándolo en el mismo momento en que se produce. Y proporcionando las herramientas de gestión y análisis forense necesarias para poder recuperar los importes, cuando se ha producido.

Más aún, la tecnología de prevención del fraude permite incrementar las cifras de negocio no sólo reduciendo el dinero defraudado, sino permitiendo operaciones de negocio (a un menor riesgo), que por los medios tradicionales quedarían desestimadas. Y más operaciones implican más negocio y mejores resultados.

**Ángel Gavín, Project Manager  
de GMV Secure e-Solutions**

## Tecnología para la detección de amenazas y análisis forense

GMV tiene una amplia experiencia desplegando, configurando y operando SIEM (Security Information and Event Management) propios y en clientes. Eso les ha llevado a detectar los mayores problemas a los que se enfrentan los SIEM tradicionales; los escasos medios de personal para que opere y actúe adecuadamente (por no hablar de la llamada “fatiga de alarmas”), las limitaciones de las reglas de experto frente a ataques desconocidos, los problemas de escalabilidad y la ineficiencia de entornos con varios SIEM de distintos fabricantes.

La propuesta de GMV consiste en “aumentar” las capacidades de los SIEM para vencer las dificultades anteriores. Gracias a Big Data y la Inteligencia Artificial “estamos revolucionando las capacidades de los SIEM con lo que denominamos SIEM de Nueva Generación (SIEM NG)” señalan sus expertos.

No se trata “de reemplazar los SIEM actuales, sino complementarlos con una arquitectura Big Data y capacidades de Inteligencia Artificial usando soluciones Open Source debidamente integradas”. De esta forma, “creamos verticales personalizados que integran con cualquier SIEM, y que se alimentan de las fuentes de datos externas e internas que los SIEM tradicionales no suelen considerar (por ejemplo, sondas específicas)”. Más aún, los SIEM NG “aprenden como un operador de SIEM más, priorizando alertas y eventos, reduciendo la fatiga de alarmas y permitiendo una explotación del SIEM más eficiente y económica con menos recursos”.

Un caso de uso de estos nuevos SIEM incluye la vigilancia de actividad de antiguos empleados, que pretendan (por ejemplo) el acceso a servicios intencio-



*Gracias a Big Data  
y la Inteligencia Artificial  
estamos revolucionando las  
capacidades de los SIEM  
con lo que denominamos SIEM  
de Nueva Generación*

nadamente o no. Se estudian y clasifican los usuarios en categorías, según cuál sea su interacción con los sistemas IT de la empresa. La exfiltración de información en ataques avanzados, o la detección de eventos complejos (tal y como el intento de fraude por reseteo de contraseña) son otros ya implementados. Como caso avanzado de análisis sobre estos SIEM NG la compañía reseña aquel en el que se distingue entre usuarios humanos y malware avanzado en conexiones, como Dridex. Se trata de un malware especializado en el robo de credenciales bancarias que ejecuta ataques de tipo mando y control (C2). Para combatir este troyano, GMV desarrolló un módulo para el SIEM NG basado en un ensamblado de 6 algoritmos de Machine Learning que se alimentaba de los logs de conexión de los usuarios y definía patrones de comportamiento de éstos mediante sus sesiones, resultando en un éxito rotundo.

Ángel Gavín, Project Manager  
de GMV Secure e-Solutions

## Participación de forma activa en el nuevo reto de la Industria 4.0

En la actualidad las empresas necesitan dar un cambio estratégico histórico apostando por la investigación, el desarrollo y la innovación, así como la incorporación de nuevas tecnologías en productos y procesos de fabricación, y la mejora de sus niveles de calidad, seguridad, fiabilidad y reducción de impacto ambiental, en relación

Está claro, por tanto, que hay grandes oportunidades en el desarrollo de nuevos negocios que no solo van dirigidas al uso de técnicas para la optimización de la vida de las máquinas optimizando repuestos sino también por la necesidad de integrar estas tecnologías en los nuevos sistemas.

*La modelización y simulación de los procesos industriales para culminar con la optimización de los mismos es un claro ejemplo de la aplicabilidad de clásicas herramientas estadísticas y de novedosos algoritmos*

con los sectores industriales. La modelización y simulación de los procesos industriales para culminar con la optimización de los mismos es un claro ejemplo de la aplicabilidad de clásicas herramientas estadísticas y de novedosos algoritmos de Machine Learning y Deep Learning en los cuales GMV cuenta con dilatada experiencia. El empleo de herramientas Open Source, así como del uso de servicios en la nube, acelera de forma sustancial la implementación de dichas soluciones Big Data reduciendo, al mismo tiempo, la inversión económica necesaria para su introducción y posterior empleo.

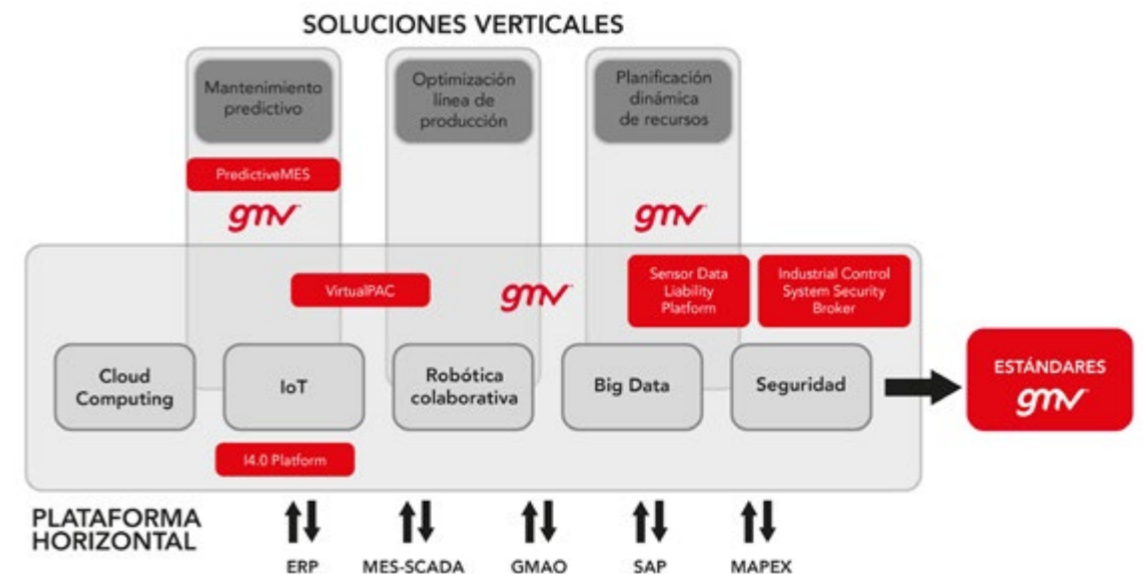
Por otro lado y de acuerdo a las últimas previsiones, el mercado alrededor del mantenimiento predictivo llegará a los 10.000 millones de EUR en el año 2022 sobre todo debido a las tecnologías de Internet de las Cosas (IoT).

En esta misma línea, GMV lidera [PRODUCTIO](#) (“PROductivity InDUstrial EnhancEment through enabling TechnOLOGies”) un proyecto del CDTI, centrado en la investigación sobre nuevas tecnologías para la mejora de los procesos de mantenimiento industrial, que ayuden a predecir las anomalías y fallos, reduciendo los tiempos de parada y aumentando la disponibilidad de las máquinas. Tecnologías como técnicas de Inteligencia Artificial en mantenimiento predictivo para poder monitorizar y mantener máquinas distribui-

das por el mundo; tecnologías novedosas de fabricación relacionadas con la industria conectada que apoyen a la toma de decisiones en la fase de producción y mantenimiento; predicción de fallos en sistemas y fórmulas para mejorar el nivel de eficiencia global del proceso industrial utilizando herramientas analíticas (Big Data).

En definitiva, obtener conocimiento para convertir máquinas y herramientas en sistemas ciberfísicos capaces de mejorar aspectos de fiabilidad, rendimiento, disponibilidad, productividad y calidad; y en tecnologías que permitan asegurar la integridad de los datos de sensores y evitar el uso fraudulento por parte de los usuarios.

**Pedro José Hernández,**  
**Industry Big Data & Digitalization**  
**Senior Consultant de GMV Secure e-Solutions**



## El Big Data está revolucionando el sector sanitario

El uso de las tecnologías de la información y comunicaciones en el ámbito de la salud ha permitido disponer, a día de hoy, de ingentes cantidades de datos procedentes de fuentes tales como: la historia clínica electrónica, radiología, asistencia domiciliaria, laboratorio y un sinfín más de microsistemas, que, junto con la tendencia a la interoperabilidad e integración, así como el avance de las redes de comunicaciones ofrecen un entorno propicio para hacer realidad la Digitalización Sanitaria. En este contexto se hace viable desarrollar productos y soluciones que ayudan a disponer de información de valor en tiempo y con un equilibrio coste-efectividad.

La necesidad de disponer de información de evidencia para conducir y personalizar los procesos asistenciales ha derivado en la creación de nuevos modelos de análisis masivos de datos (Big Data) asociados a cada servicio de salud, con capacidad de ser comparados y relacionados, de obtener medidores de eficiencia y estimadores de riesgo. En definitiva, centrarse en el pa-

*La necesidad de disponer de información de evidencia para conducir y personalizar los procesos asistenciales ha derivado en la creación de nuevos modelos de análisis masivos de datos (Big Data)*

ciente y su entorno socio-sanitario para una mejora en los resultados de salud.

Así pues, desde la tecnología dos son los elementos clave que pueden contribuir a mejorar estos resultados: Datos y Evidencia. Y son la ciudadanía, la asistencia clínica, los estudios epidemiológicos, la investigación clínica, la industria farmacéutica... las distintas organizaciones proveedores de servicios de salud las máximas donantes de datos (Real World Data) y, a su vez, las mayores consumidoras de conocimiento (Real World Evidence).

Aplicando tecnología Big Data se optimiza el trabajo de análisis de grandes cantidades de datos. En este contexto, el uso de esta tecnología permitirá caracterizar más específicamente las distintas enfermedades y sus grupos de riesgo. Esta estratificación implica identificar marcadores (ya sean clínicos o moleculares) que definan a la población afectada por una determinada enfermedad con el objetivo de realizar diagnós-



tics más precisos y desarrollar tratamientos dirigidos y adaptados al riesgo de cada persona. De este modo, será posible “customizar” el tratamiento a las características de cada persona, administrando terapias dirigidas y evitando la exposición a toxicidades innecesarias.

A la hora de trabajar en proyectos de esta naturaleza, GMV se adecua a los protocolos que en materia de Seguridad y Privacidad dictan las autoridades competentes y que regularizan la forma de extracción y uso que se dará a los datos, así como en el desarrollo e implantación de herramientas sofisticadas, bajo tecnología Big Data y algoritmos matemáticos, para la explotación de los datos y obtención de evidencia a través de los procesos de ingesta y calidad, estandarización, modelización, enriquecimiento y analítica descriptiva y predictiva.

**Inmaculada Pérez,**  
**Responsable Proyectos de Big Data**  
**en Salud de GMV Secure e-Solutions**

## Chatbots, los asistentes que nos atenderán en el futuro

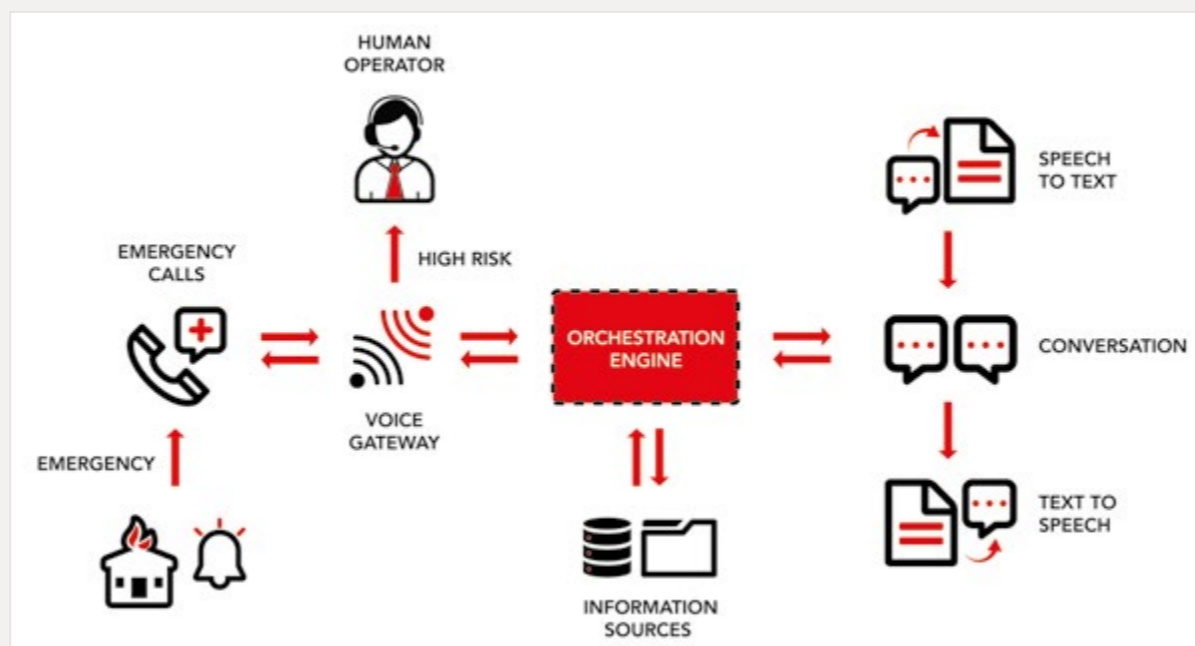
Los chatbots son hoy en día la aplicación de consumo número uno de inteligencia artificial y de la tecnología cognitiva. Según TechEmergence se espera que durante los próximos cinco años siga siendo así. Muchos expertos llamaron a 2016 “el año de los chatbots”, porque miles de ellos ya ayudan a las empresas a mejorar el servicio al cliente, vender más y aumentar sus ganancias.

Gartner también predice que, en 2020, más del 85% de las interacciones con los clientes se gestionarán sin intervención humana. La importancia que los gigantes tecnológicos como Google, Facebook, Microsoft, IBM y Amazon están dando a las tecnologías cognitivas, y más concretamente a los chatbots, es un fuerte indicador de que jugará un papel clave en el futuro. El equipo de Big Data y Business Analytics de GMV, siempre vigilante ante nuevos retos y oportunidades, aceptó el desafío que propuso el concurso internacional IBM Watson Build Challenge, cuyo objetivo era crear aplicaciones con la ayuda de la plataforma IBM Bluemix y la tecnología cognitiva IBM Watson. La exitosa propuesta de GMV es un asistente para la gestión de emergencias conocida como **HelpBot**.

Estamos acostumbrados a ver chatbots en entornos comerciales, algunas veces incluso con fines poco éticos, pero **HelpBot** pone la tecnología al servicio de las

personas, con un fin humanitario como es el de ayudar a las autoridades a resolver los principales problemas a los que se enfrentan en situaciones de emergencia.

En este tipo de situaciones, el número de llamadas crece exponencialmente y las autoridades adolecen de los recursos necesarios para poder atender todas de



foma ágil. **HelpBot** resuelve esta situación de “colapso” proporcionando una gestión elástica de los recursos, activándolos cuando sea necesario, y aumentando la capacidad de respuesta de las centralitas de emergencias convencionales.

Otra situación que caracteriza a las emergencias es que las autoridades se encuentran con dificultades a la hora de priorizar las llamadas en función de su naturaleza, o de los criterios especificados en el Plan de Emer-

gencia del organismo correspondiente. En este tipo de situaciones se reciben numerosas llamadas para solicitar información, ofrecer ayuda, o interesarse por sus familiares. Si las llamadas no se priorizan convenientemente pueden provocar retrasos en los rescates, y en el peor de los casos derivar en problemas más serios como pérdidas de vidas humanas.

**HelpBot** es una herramienta inteligente para ayudar a la gestión eficaz de las emergencias en la fase de la comunicación con afectados y familiares. Cuenta con capacidad para reconocer la voz del usuario, detectar su intención, y en una situación de alto estrés o peligro, redirigir la llamada a un operador humano. Las llamadas menos prioritarias serán atendidas por **HelpBot**, liberando de este modo a los operadores humanos para que puedan centrarse en atender a las víctimas. Gracias

a su alta disponibilidad, **HelpBot** también puede dar respuesta a una situación frecuente en las centralitas de emergencias: alta rotación y bajas del personal por estrés y depresión tras una catástrofe o desastre, cifra que supera hasta el 60%.

**Inmaculada Perea,**  
**Data Scientist de GMV**  
**Secure e-Solutions**



## La inteligencia artificial ya aprende por su cuenta

**Hace unos meses pudimos presenciar como DeepMind presentaba la última versión de su software para jugar al juego de mesa oriental AlphaGo Zero. Este programa de Inteligencia Artificial fue presentado en 2015 por la empresa de Google y tiene como objetivo dominar el juego de mesa oriental Go basándose en técnicas de aprendizaje por refuerzo, redes neuronales y árboles de búsqueda.**

En 2016, como algunos podrán recordar, se produjo la primera partida oficial de la primera versión de AlphaGo contra el profesional Fan Hui quien fue vencido 5 juegos a 0 y del cual cogería el nombre para dicha versión.

Desde ese momento el programa fue evolucionando a través de distintas mejoras tanto a nivel algorítmico como de hardware permitiéndole vencer en mayo de 2017 a Ke Jie, considerado mejor jugador de Go del mundo hasta el momento, en los tres juegos que se disputaban. Esta

versión del programa, AlphaGo Master, demostró la superioridad de la máquina sobre la habilidad humana en dicho juego.

Así pues, la versión AlphaGo Zero fue presentada a través de un paper en Nature. A diferencia de las versiones anteriores de AlphaGo, que se entrenaban a través de miles de estrategias y jugadas planteadas por humanos, AlphaGo Zero se desprende de esta parte y realiza el entrenamiento contra sí mismo a través de jugadas iniciadas de forma aleatoria teniendo como única información las reglas básicas del juego. Esta técnica es más fuerte que el resto de versiones ya que no está limitada por el conocimiento humano. En cambio, es capaz de aprender sin conocimientos anteriores a partir de su propia experiencia convirtiéndose en el mejor jugador del mundo de Go en tan solo 40 días de entrenamiento.

Como ha descrito el equipo de este proyecto, AlphaGo Zero es capaz de realizar el entrenamiento gracias a una nueva técnica de aprendizaje por refuerzo basada en el uso de una red neuronal sin conocimiento del juego que empieza a jugar contra sí misma en combinación con su algoritmo de búsqueda. A medida que juega, la red neuronal se reajusta y actualiza para predecir los movimientos y el futuro ganador de la partida. Esta actualización procesa en cada iteración una nueva versión de AlphaGo

Zero más fuerte mejorando el rendimiento y la calidad de juego.

Aparte de ser mejor jugador, encontramos otras mejoras respecto a sus versiones anteriores. Una de ellas es la mayor flexibilidad que ofrece gracias al poco consumo computacional que el programa requiere. AlphaGo Zero es capaz de ejecutarse en tan sólo 4 TPU mientras que las versiones anteriores usaban 48.

Por otro lado, al ser un programa autodidacta, este programa nos demuestra que es posible desarrollar nuevos algoritmos mucho más avanzados sin depender de una cantidad ingente de datos. Esto era algo bastante criticado por los expertos en el campo de la inteligencia artificial ya que para conseguir resultados óptimos era necesario el uso masivo de datos y un alto gasto computacional.

¿Cómo de útil puede llegar a ser esto de cara al futuro? La respuesta la encontramos extrapolando esta inteligencia para buscar resultados complejos. Al no estar programado específicamente para comprender Go, este podría reprogramarse para estudiar la información en otros campos de investigación como: descubrimiento de fármacos, diseño de materiales, química cuántica...

**Gabriel Escudero, Data Scientist  
de GMV Secure e-Solutions**



¿Te ha gustado este reportaje?

Compártelo en tus redes sociales



El futuro sanitario está ligado a extraer valor de los grandes volúmenes de datos de que dispone

No cabe duda de que la Sanidad es un sector capaz de ver mejorar sus resultados introduciendo tecnologías de Big Data en su actividad. A día de hoy, ya se están realizando numerosos proyectos relacionados con la recopilación, gestión y explotación de datos clínicos y epidemiológicos que persiguen extraer información de calidad y útil para que los especialistas practiquen una medicina basada en la evidencia. Entre algunos en los que GMV está presente podemos mencionar, HARMONY, PAPHOS o MOPEAD.

Ciertamente, las capacidades de la tecnología Big Data es transversal a la mayoría de los sectores. La agricultura de precisión por observación de la tierra e IoT es otro campo en el que GMV también se ha introducido con su desarrollo [wineo](#): un sistema de precisión basado en imagen por satélite y Data Mining que permite agilizar la toma de decisiones. [wineo](#) utiliza datos meteorológicos que mejoran la toma de decisión en aspectos tan relevantes para la agricultura como son la necesidad de regadío o momento de recolecta, reduciendo el consumo de agua, pesticidas y fertilizantes, minimizando tanto la huella de carbono, como el impacto de pesticidas en los produc-



tos agrícolas. El sistema estima una reducción de costes directos de producción de entre 4 y el 6 por ciento, que conlleva una mejora sustancial de los beneficios netos de la producción agrícola.

Otros ámbitos en los que GMV está aplicando las tecnologías que vienen de la mano de la Era Digital es en la detección de amenazas y análisis forense; la prevención del fraude empleando Inteligencia Artificial; la detección de anomalías en centros de datos, la optimización de procesos industriales e IoT, o AlphaGo Zero, un programa de Inteligencia Artificial que tiene como objetivo dominar el juego de mesa oriental Go basándose en técnicas de aprendizaje por refuerzo, redes neuronales y árboles de búsqueda.

Y si bien hoy toda organización está en la red, es importante que éstas conozcan la información que se maneja allí sobre ellas y disponer de medidas capaces de evitar situaciones de riesgo reputacional. Para ello GMV ha diseñado el servicio de vigilancia digital, [atalaya](#), con el que GMV monitoriza diversas fuentes, en modalidades de 8x5 o 7x24, con el objetivo de identificar informa-

ción que permita detectar publicaciones accidentales o premeditadas que puedan suponer un riesgo a la organización o un perjuicio a su imagen. Actividades como fuga de información, suplantación de identidad, ataques organizados, constatación de actividades fraudulentas, espionaje industrial, ATP, phishing, pharming, spam, información obtenida a través de ataques contra la organización, credenciales de usuarios en los diferentes servicios de la organización... son detectadas por este servicio de vigilancia.



#### Enlaces relacionados



[Casos de éxito Big Data](#)



[GMV](#)



[GMV News](#)