



Soluciones Data & IA

# PREDICCIÓN DE INGRESO UCI EN PACIENTES CON COVID-19

<b>CLIENTE</b>	Ribera Salud
<b>PROYECTO</b>	Predicción de ingreso en UCI en pacientes con Covid-19
<b>SECTOR</b>	Salud
<b>SEDE</b>	Valencia
<b>SOLUCIÓN</b>	Machine Learning

## 01 Grupo empresarial Ribera Salud

Ribera es un grupo empresarial que desde 1997 es proveedor de servicios sanitarios públicos y privados. Con más de 7.000 profesionales, su vocación es transformar la manera de trabajar e innovar en la gestión sanitaria. Ha desarrollado un modelo sanitario que se estudia cómo caso de éxito en reconocidas universidades internacionales como Harvard Business School o Berkeley.

El grupo gestiona varios proyectos hospitalarios y servicios asistenciales especializados en España, Latinoamérica y Europa. Además, cuenta con una Escuela Universitaria de Enfermería y Centro de Estudios de Formación Profesional Sanitaria, una central de compras: Ribera B2B; una división de laboratorios: Ribera Lab y una empresa tecnológica: Futurs.

Ribera es una empresa de salud responsable, que trabaja para mejorar la salud y el bienestar de los ciudadanos. Su compromiso social queda patente a través de múltiples iniciativas y líneas de trabajo, que acorde a los Objetivos de Desarrollo Sostenible y a la Agenda 2030, aportan valor a la sociedad actual y futura y contribuye a la sostenibilidad de los sistemas sanitarios.

## 02 Inteligencia Artificial en la sanidad

la Inteligencia Artificial (IA) está presente en el descubrimiento de nuevos fármacos, así como en la interpretación de imágenes radiológicas, y hasta en el análisis del genoma de un paciente, ayudando a entender la progresión de una enfermedad. Además, optimiza los recursos sanitarios disponibles, automatizando tareas repetitivas, y acelerando la interpretación de los datos.

Y es que, con la proliferación de registros de salud electrónicos, intercambios de información sanitaria, y la información procedente de los dispositivos wearables y sensores, se dispone cada vez de mayor cantidad de datos que, sin ayuda de la IA, serían imposibles de analizar y procesar.



### RIBERA SALUD APUESTA POR LA IA

En Ribera Salud surge la oportunidad de aplicar técnicas avanzadas de IA para detectar el empeoramiento clínico (ingreso en UCI) de pacientes con diagnóstico asociado a la COVID-19. De este modo, se le brinda al médico la oportunidad de replantearse el plan terapéutico del paciente y al gestor le facilita la predicción de la ocupación de camas UCI.

## 03 Covid-19 y ocupación UCI

En diciembre del 2019 irrumpió una nueva neumonía causada por la COVID-19. Según los datos hasta diciembre de 2021, el número de contagios por COVID-19 es superior a 250 millones de personas y de 5 millones de fallecimientos. En España los casos de contagios eran de 5,17 millones y cerca de 90.000 fallecimientos.

Esta situación supuso un **reto mundial para los sistemas sanitarios**, por su elevado consumo de recursos y su alto índice de gravedad.

Las altas tasas de hospitalización y el número limitado de camas de UCI, tuvo implicaciones importantes a nivel logístico y ético. La saturación de estas unidades puso en riesgo la respuesta del sistema sanitario a la enfermedad de la COVID-19.

*“Las técnicas de IA han sido ampliamente utilizadas para el cribado, el diagnóstico, y para predecir el pronóstico de paciente Covid-19 basados en imágenes radiológicas y en datos clínicos, de laboratorio y demográficos. Muy pocos se centran solamente en el ingreso en UCI o necesidad de ventilación mecánica invasiva”.*



## 04 Necesidades de Ribera Salud

Ribera Salud tenía desarrollada una **solución de Machine Learning** para ayudar a predecir el hecho de que un paciente fuese a entrar en UCI por COVID-19.

**La solución ML tenía dos objetivos:**

- Facilitar la toma de decisiones clínicas
- Ayudar a la gestión de recursos escasos

Sin embargo, necesitaban ayuda experta para revisar que el trabajo realizado seguía las prácticas adecuadas del desarrollo de este tipo de proyectos de Machine Learning. Además, tenían la necesidad de poner en producción dicho modelo de una forma automatizada.

## 05 Objetivos del proyecto

### 1. Revisión del modelo actual y conjunto de datos.

El primer objetivo fue validar el método de creación de su modelo de Machine Learning, de cara a buscar posibles mejoras con las que enriquecer el modelo y mejorar su rendimiento. Por otro lado, también debíamos analizar detalladamente el conjunto de datos proporcionado para buscar posibles anomalías o datos sin tratar que puedan estar empeorando el rendimiento del modelo.

### 2. Enriquecimiento del modelo.

Como segundo objetivo, tras la exploración exhaustiva de los datos y la revisión del modelo actual, se planteó dar un paso más allá y desarrollar una solución más completa y vanguardista. La aplicación de sofisticadas técnicas para el tratamiento de los datos y la creación de nuevas variables (conocido como feature engineering), así como la prueba de diferentes algoritmos permitió obtener un modelo ML óptimo y generalizable.

### 3. Despliegue y automatización.

El tercer y último objetivo era desplegar la solución ML en la plataforma del cliente. Un modelo puesto en producción en un entorno, como puede ser Azure, que nos permitiese automatizar el mantenimiento y los reentrenamientos del modelo, así como el consumo del mismo. La implementación de CI/CD permitiría integrar de forma automática los cambios que se produzcan en el código y actualizar en consecuencia el modelo en producción.

## 06 Verne ML Framework

Para llevar a cabo el análisis de los datos y el desarrollo del modelo de Machine Learning se hizo uso del “**VerneML framework**”. Una herramienta desarrollada por el equipo de Machine Learning de Verne Tech que permite conseguir resultados “en el acto”.

Es un framework multipropósito, con el cual abarcamos la resolución de diversos casos de uso clásicos. Cuenta con una configuración estándar que permite obtener rápidamente un resultado baseline y saber “cómo es la situación”, ofreciendo la suficiente flexibilidad para customizar y refinar los modelos para mejorar los resultados en nuevas iteraciones.

## 07 Solución implementada

### 1. Análisis de Datos

El análisis exploratorio de datos (EDA) es la práctica para describir los datos por medios de técnicas estadísticas y de visualización para enfocar aspectos importantes dentro de los datos para un análisis posterior. Esto implica mirar un conjunto de datos desde muchos ángulos, describirlo y resumirlo sin hacer suposiciones sobre su contenido. Este paso es imprescindible antes de sumergirse en la creación del modelo, entre otras cosas para asegurarse de que los datos sean realmente lo que creemos que son.

La muestra total fue de casi 2.000 pacientes y un centenar de variables catalogadas como factores predictores de empeoramiento clínico en pacientes con COVID-19. Se trata un volumen de datos muy pequeño, en términos de ML, que hace más costoso la obtención de un buen modelo predictivo.

Para comenzar se analizaron en profundidad las variables, sus distribuciones y su composición. A partir de aquí, se preprocesaron los datos, se descartaron algunas variables e imputaron los valores nulos con valores de negocio (con sentido sanitario en este caso) y se preparó un nuevo conjunto de datos con las características más importantes.

Hubo múltiples iteraciones con Ribera Salud para eliminar variables desactualizadas y problemáticas (por ejemplo, medicamentos contra la COVID-19 que ya no recomendaba la OMS y dejaron de usarse).

## 2. Mejora del modelo.

La solución para el segundo objetivo se enmarca en un problema de tipo Aprendizaje Supervisado en Machine Learning. La predicción de si un paciente requerirá ingreso en UCI es un valor binario, por lo que utilizamos un modelo de Clasificación.

El rendimiento de cuatro de los algoritmos más usados en problemas de clasificación fue evaluado de acuerdo con el objetivo:

- LightGBM
- XGBoost
- Logistic Regression
- Random Forest

La métrica utilizada para seleccionar el mejor algoritmo fue el recall, también conocido como sensitivity (SN), ya que maximizando esta métrica se reduce el número de falsos negativos. El recall es la métrica por excelencia cuando se trata de la predicción de diagnóstico médico en pacientes. Es de importancia vital minimizar que un modelo prediga una enfermedad o un empeoramiento clínico cuando no lo hay, hablamos de personas. Se seleccionó el algoritmo con las mejores métricas de clasificación. Una vez en este punto se optimizó este modelo usando técnicas avanzadas de ML para lidiar con el desbalanceo de la variable a predecir (12% de pacientes UCI frente al 88% no UCI) y reducir posibles sesgos en las predicciones.

## 3. Interpretabilidad e Fairness

Dada la sensibilidad del estudio, como parte de la mejora del modelo, se incluyó un módulo de interpretabilidad del modelo. Interpretar un modelo nos ayuda a entender mejor los resultados y saber qué variables influyen en las predicciones a nivel global.

Esta interpretación, además, nos indica qué valores, más altos o bajos, indican una mayor probabilidad de UCI para cada uno de los pacientes (nivel local). Esta herramienta permite analizar detalladamente los posibles fallos que cometa el modelo en sus predicciones y saber si se trata de un caso anómalo (por ejemplo, un paciente que empeora de una forma rápida e inesperada) o es un verdadero error del modelo.

Esto permite validar el modelo de una forma más fiable por parte de Ribera salud y analizar los falsos positivos y negativos en detalle.

Dentro de este módulo incluimos también una herramienta para medir la equidad (más conocido como fairness en inglés) del modelo en sus predicciones para distintos rangos de edad. La edad es un indicador fundamental para el diagnóstico y tratamiento de enfermedades. Se crearon, de la mano de Ribera Salud, los rangos de edad pertinentes y se validó que las predicciones del modelo fueran equitativas para todos ellos.

Esta técnica permite mirar con lupa el modelo ML y depurar posibles sesgos que aparezcan con el paso del tiempo.

#### 4. Arquitectura de la solución.

La solución se implementó con tecnologías Azure de Microsoft, siendo este el entorno habitual de trabajo del cliente. El análisis de los datos y el desarrollo del modelo se llevaron a cabo haciendo uso del **framework ML** propio de **Verne** basado en el lenguaje de programación **Python**.

Para el despliegue del código, se usaron contenedores Docker orquestados con **Azure Kubernetes Service**. Esta tecnología ofrece una serie de ventajas imprescindibles para cubrir las necesidades del cliente:

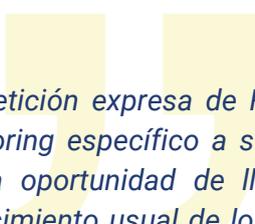
- **Estabilidad en la solución:** Evitar que la caída del contenedor provoque intervención manual y cambios de contraseña.
- **Escalabilidad:** Bajo situación de demanda el servicio no se verá afectado.
- **Seguridad:**
  - Tráfico cifrado: Uso de SSL.
  - Solución protegida por contraseña.
  - Filtrado de IP para acceso a la solución.
- **Mantenimiento:** Permite estrategias de despliegue continuo con actualización sin pérdida de servicio y recuperación rápida en caso de errores en una nueva versión.

Además, se incluyeron estrategias de automatización en la compilación y despliegue de la solución usando **Azure DevOps** para implementar CI/CD.

#### 5. Transferencia de conocimiento y formación

Durante el proyecto hubo un parte notable de tiempo dedicada a la formación del equipo técnico del cliente. En particular, se formó a un perfil Data Scientist sobre las tecnologías y técnicas ML utilizadas y dotarle, de este modo, de las herramientas necesarias para el mantenimiento y reentrenamiento del modelo desarrollado.

Como en todo proyecto, se dedicaron varias sesiones a la transferencia de conocimiento. En estas sesiones se validó el modelo con el equipo de negocio y se explicó detalladamente la arquitectura de la solución al equipo técnico.



*"La petición expresa de Ribera Salud de hacer un mentoring específico a su científico de datos nos dio la oportunidad de llevar la transferencia de conocimiento usual de los proyectos un paso más allá. Dentro del marco del proyecto, pudimos dar una formación técnica avanzada que encaja a la perfección con el ADN de Verne y que marca la diferencia en los servicios que ofrecemos"*

*Ignacio García  
Data & AI Architect de Verne Group*

# 08 Resultados obtenidos

## Módulo predictivo.

Los resultados obtenidos superaron el 90% en las todas las métricas de clasificación (precisión, sensibilidad, AUC y accuracy). Lo cual, son resultados excelentes.

## Publicación científica.

Dado que Ribera Salud es un grupo empresarial que apuesta por la tecnología y la innovación, posee publicaciones en importantes revistas científicas del sector de la salud. De la solución desarrollada en este proyecto, saldrá también una publicación científica, en la cual Verne ha sido partícipe y aparecerá en la lista de autores como entidad colaboradora.

## Situación actual.

Al finalizar el proyecto, cambiaron su modelo Machine Learning antiguo y pusieron el nuevo desarrollado en el entorno de producción.

Tras pasar dos semanas, el feedback recibido por Ribera Salud fue positivo, el modelo estaba funcionando como se esperaba, resolviendo los objetivos inicialmente planteados. A día de hoy, el cliente ha mostrado independencia total para mantener y consumir la solución desarrollada, tanto a nivel de arquitectura en Azure como el modelo ML.

La buena sintonía con el cliente y la profesionalidad mostrada por ambas partes abre la puerta a futuras colaboraciones en nuevos proyectos.

*“Ha sido muy gratificante poner el Machine Learning al servicio de la sanidad de la mano de Ribera Salud. Hemos conseguido desarrollar un modelo capaz de predecir qué pacientes con diagnóstico COVID-19 requerirán atención en UCI con un alto nivel de precisión. La principal clave del éxito ha sido la estrecha colaboración y la transferencia de conocimiento bidireccional con Ribera Salud, tanto con su científico de datos como con el equipo de negocio. Gracias a ello, el proyecto ha sido un éxito y hemos dotado a los sanitarios de un herramienta para optimizar la ocupación de camas UCI en tiempos de pandemia.”*

*Ignacio García  
Data & AI Architect de Verne Group*