
This is the **published version** of the master thesis:

Alberto Cadavid, Carlos; Lozano, Antonio , tut. Evaluación del desempeño de modelos de Aprendizaje Automático para gestión del recurso hospitalario. 2025. 18 pag. (Màster en Intel·ligència Artificial i Big Data en Salut Diplomats en Educació Social)

This version is available at <https://ddd.uab.cat/record/318698>

under the terms of the  license

Evaluación del desempeño de modelos de Aprendizaje Automático para gestión del recurso hospitalario.

Carlos Alberto Cadavid.

TDF Inteligencia artificial en salud.

Universidad Autónoma de Barcelona.

Resumen: La inteligencia artificial se ha convertido en una herramienta transformadora en la gestión de la salud, facilitando el desarrollo de sistemas de apoyo para la toma de decisiones basadas en datos en tiempo real. Su potencial trasciende al ámbito clínico e impactando también la gestión administrativa. Este proyecto evalúa diversos modelos de aprendizaje automático (Machine Learning (ML)) aplicados a una base de datos hospitalaria. Con el objetivo de identificar patrones históricos que permitan estimar la demanda de recursos y así, optimizar el uso del talento humano, la tecnología, reducir costos y mejorar los estándares de calidad.

Palabras claves: Inteligencia artificial, aprendizaje automático, administración en salud, modelos predictivos, eficiencia operacional, sistemas de apoyo en la toma de decisiones.

Abstract: Artificial intelligence has become a transformative tool in healthcare management, enabling the development of real-time, data-driven decision support systems. Its potential spans both clinical and administrative domains. This project evaluates various machine learning models applied to a hospital dataset to extract historical patterns that support forecasting healthcare resource demand, ensuring optimal use of human and technological assets, cost reduction, and quality improvement.

Keywords: Artificial Intelligence, Machine Learning, Healthcare Administration, Predictive Models, Operational Efficiency, Decision Support Systems.

- *E-mail de contacto: cadavid9@hotmail.com*
- *Trabajo tutorizado por: Antonio Lozano*
- *Curso: 2025*

1.0 INTRODUCCIÓN

Los sistemas de salud están experimentando demandas sustanciales en sus capacidades y recursos hospitalarios, esto debido al incremento en el volumen de pacientes, a la complejidad de las enfermedades y a una población más envejecida y con una mayor carga de enfermedad. Lo que ha llevado a un desborde en el uso de los recursos en salud, que hace inviable cualquier sistema sanitario.

Tradicionalmente el manejo de los recursos se ha hecho con base en datos históricos y en la experiencia del personal administrativo, la aplicación de inteligencia artificial (IA) en salud se convierte en una oportunidad con soluciones que posiblemente permitirán transformar la manera de administrar los recursos sanitarios.

Los modelos de aprendizaje automático se presentan como una solución que permite la toma de decisiones basadas en datos en tiempo real. Es importante destacar que el 80% de los datos son no estructurados o son pseudo-estructurados, para lo cual las opciones como los procesadores de lenguaje natural (PLN) ampliaron la capacidad de captura, organización, integración y análisis (1).

Este proyecto, pretende hacer un análisis exploratorio sobre el desempeño de modelos de Aprendizaje Automático, evaluando su capacidad para la estimación de la estancia, egresos mes, costos y estado al alta, como variables determinantes para la demanda de los recursos sanitarios.

“Se plantea la hipótesis. Los modelos de Aprendizaje Automático, debidamente entrenados con datos hospitalarios estructurados, permiten estimar de manera suficientemente precisa, eventos como estancia hospitalaria, costos, egresos mes y sobrevida.”

2.0 ESTADO DEL ARTE

¿La IA y sus modelos de aprendizaje automático llegarán a ser la solución para la distribución transformativa del recurso? ¿Permitirán anticipar la demanda de servicios, el manejo óptimo de la cadena de suministros, y la asignación del recurso humano y tecnológico? (2).

Los sistemas de salud están basados en la administración de los datos históricos y en la experiencia de las personas a cargo, lo que obligatoriamente impone una carga de empirismo en la toma de decisiones. La IA pone a nuestra disposición, nuevas formas para administrar basada en datos y en tiempo real. Permitiendo identificar patrones y correlaciones que pueden no ser aparentes al ojo humano a través del análisis tradicional, esta capacidad es particularmente valiosa en casos complejos donde múltiples factores contribuyen al resultado de los procesos de atención en salud (2, 6).

No está exenta de retos y limitaciones, relacionados con la privacidad y custodia de datos sensibles, como también evitar el sesgo, buscando la mayor representatividad posible y asegurando una información de buena calidad. La interoperabilidad y el desarrollo óptimo de aplicaciones de caja blanca son otros dos aspectos necesarios en la implementación que permitirá generar confianza y valor en la práctica clínica, igualmente, uno de los debates más candentes de la IA en salud está centrado en la confiabilidad, ¿pueden las predicciones (futuro) con datos del pasado y en especial en un mundo dinámico y cambiante, asegurar una información veraz?, esto posiblemente subyugará la IA a una fuente permanente de datos vigentes (2).

Los modelos predictivos utilizan métodos matemáticos computacionales para predecir eventos futuros, basados en técnicas de simulación. La aplicación de estas, en el dominio de la salud es retardadora, debido a la naturaleza dinámica y compleja de los datos, sumado al hecho que la información en los registros electrónicos, en su mayoría son datos no estructurados, lo que obliga a cumplir con unos mínimos de volumen, máximos de valor y veracidad que soporten el nivel de certeza dichas predicciones.

Principios claves del modelamiento predictivo:

Los métodos computacionales y matemáticos incluyen varias técnicas como análisis de regresión, arboles de decisión, Random Forest, técnicas de agrupación y asociación, redes neuronales, máquinas de soporte vectorial entre otros.

Podemos separar en dos grandes grupos los modelos predictivos (3):

- Modelamiento predictivo basado en ecuación.
 - Regresión lineal
 - Regresión logística
 - K-means
 - Máquinas de Soporte Vectorial
 - NavyBayes
 - Algoritmos Apriori
- Modelamiento predictivo computacional
 - Redes neuronales

Los modelos basados en ecuaciones describen la relación entre variables, evaluando como una variable independiente influye en el resultado de una dependiente, son simples en su construcción por lo que permiten manejar grandes volúmenes de datos, sus mayores limitantes están en la naturaleza compleja de las variables que influyen en el resultado de un estado de salud determinado, además si dichas variables fluctúan en el tiempo pueden conllevar a predicciones desactualizadas.

El modelamiento predictivo basado en sistemas computacionales utiliza técnicas que no son fácilmente explicables y pueden tener dificultades para la interpretación (modelos de caja negra), dependen en gran parte de técnicas de simulación. A la fecha existen una gran cantidad de modelos validados que se pueden usar en diferentes ámbitos de la salud (7).

3.0 OBJETIVOS

General:

- Explorar la capacidad estimativa de diferentes modelos de Aprendizaje Automático, en el ámbito administrativo.

Específicos:

- Explorar el desempeño de modelos de aprendizaje automático para predicción y clasificación de variables útiles en la gestión de recurso hospitalario.
- Evaluar el comportamiento de estos modelos frente a series temporales.

4.0 MATERIALES Y MÉTODOS

Para el diseño y presentación de este proyecto se utilizaron los lineamientos de la guía TRIPOD-AI (12, 13).

Se utilizó una base de datos, disponible en el departamento de informática en salud del hospital Pablo Tobón Uribe, que contiene información demográfica, clínica y administrativa de los egresos hospitalarios desde el 2014 y hace parte para la gestión hospitalaria por GRD (Grupos Relacionados del Diagnóstico) (ver gráfico 1). Herramienta usada a nivel internacional para la gestión del recurso hospitalario, tomando datos comparativos de países de América y Europa. Para este proyecto se seleccionaron los egresos hospitalarios desde enero 1 del 2022 hasta diciembre 31 del 2024, que corresponden a 64.484 registros, con 16 columnas que contiene datos demográficos, códigos

diagnósticos en codificación CIE-10, estado al alta (vivo o muerto), clasificación por gravedad y tipo de ingreso (Médico o Quirúrgico), pagador, costo final de la atención y estancia en días, para la descripción de las variables (tomado de la aplicación Alcor GRD agrupador 3.1 GRD Internacional Refinado (IR)). (Ver tabla 1).

El proyecto está diseñado para entrenar los modelos, con los datos suministrados por la institución y serán manejados bajo la modalidad de anonimización.

Gráfico 1. Composición de la base de datos.

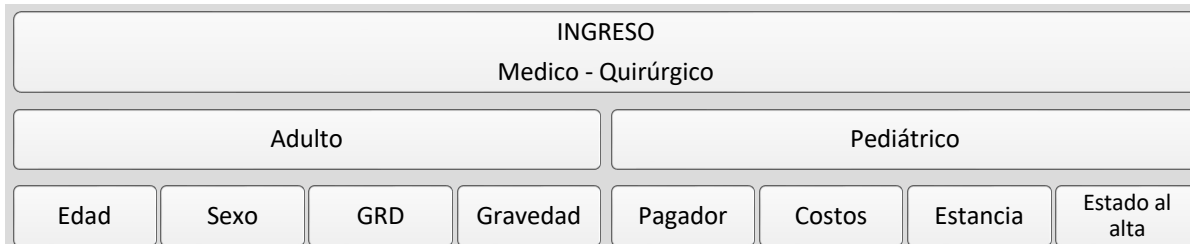


Tabla 1. (Categorización de las variables)

Variable	Tipo	Descripción
Número Consecutivo	Númerica	Orden de los registros, numero entero
Sexo	Categórica	Hombre: 1 - Mujer:2
Fecha de nacimiento	Númerica	Formato dd/mm/aaaa
Edad	Númerica	Numero entero
Fecha de ingreso	Númerica	Formato dd/mm/aaaa
Fecha de egreso	Númerica	Formato dd/mm/aaaa
Pagador (Cód)	Categórica	Código del responsable del pago
Pagador (Desc)	Categórica	Describe el responsable del pago
Motivo Egreso (Descripción)	Categórica	Vivo o Muero al alta
IR GRD descripción	Categórica	Diagnostico principal al egreso
IR Gravedad (cod)	Categórica	1: Sin complicaciones y sin comorbilidades 2: Con complicaciones y comorbilidades menores. 3: Con complicaciones y comorbilidades mayores
Costo Total	Númerica	Suma de los costos al egreso
Peso GRD Medio Inliers	Númerica	Mide la complejidad de los pacientes según el GRD, con un peso relativo >1 la casuística es de mayor complejidad.
Estancia	Númerica	Estancia de los egresos
IEMA inlier	Númerica	IEMA: Índice de Estancia Media Ajustada. Mide la eficiencia de la estancia hospitalaria, > 1 mayor ineficiencia
Estancias evitables.	Númerica	Estancias ahorradas o gastadas respecto del estándar, indica eficiencia

Para la implementación desarrollo entrenamiento y validación, se utilizó Python 3.10, en el entorno Kaggle Notebooks, utilizando las bibliotecas pandas, scikit-learn, imbalanced-learn, matplotlib y seaborn. Todo el código fuente, procesamiento y validaciones se realizaron de forma reproducible y controlada.

5.0 CONSIDERACIONES ÉTICAS Y LEGALES.

Aun con técnicas de seudonimización o anonimización existe un riesgo de re identificación. Para mitigar este riesgo se suscribe el compromiso de no re identificación por parte de los involucrados en el desarrollo. La Ley Estatutaria 1581 de 2012 establece el régimen general de protección de datos personales en la República de Colombia. La norma, por su parte, reconoce el derecho de los titulares para proteger su información y el deber de los Administradores de Datos personales, con respecto a la recolección, almacenamiento, uso, circulación de la información para finalidades constitucionales en el marco del Estado Social de Derecho.

Recientemente se publicaron los lineamientos sobre el tratamiento de datos personales en Sistemas de Inteligencia Artificial por parte de la superintendencia de industria y comercio, por medio de la circular 002 del 21 de agosto de 2024, y será esta la normatividad que rige todo lo relacionado con el tratamiento de datos. El desarrollo de los modelos y los datos utilizados por definición la hacen inclusiva y equitativa, se tomaron medidas y el máximo cuidado para que la muestra sea representativa de la población, permitiendo que el beneficio alcance a todos los usuarios (paciente, sistema de salud, desarrollador).

Una consideración final, sobre los mínimos para una herramienta informática en salud, como posición del colegio Americano de medicina (13, 14).

- Los modelos de caja negra son inaceptables.
- La complejidad y la baja usabilidad frustra el uso.
- La relevancia y el conocimiento son esenciales.
- La entrega del conocimiento y la información deben ser respetuosos

6.0 PREPROCESADO

Tabla 2. Estadística descriptiva.

Nombre	Nulos	Mín.	Máx.	Cuartil 1	Cuartil 2	Cuartil 3	Media	Desv. Estándar
Sexo (Cód)	0	1	2	1	1	2	1	1
Edad en años	0	0	107	19	49	70	45	29
IR Gravedad (cod)	0	1	3	1	2	3	2	1
Costo Total	0	600317	975017520	3490907	6745698	14492424	13545399	21935278
Peso GRD	0	0	14	1	1	2	1	2
Estancia	0	0	158	3	5	9	8	9
IEMA inlier	0	0	6	0	1	1	1	1

Basado en los hallazgos del análisis exploratorio de la base de datos (Tabla 2), se hicieron las siguientes intervenciones. No se encontraron valores nulos (NaN) por lo que no fue necesario ningún tipo de imputación, con respecto a los outliers se tuvo en consideración que los registros de estancia y costo total corresponden a datos reales y relevantes para el procesado y análisis, por tanto se hizo un tratamiento con transformación Raíz Cuadrada con la intención de reducir la dispersión sin modificar el efecto de las variables.

Se eliminaron las columnas: “Número Consecutivo”, “Fecha de nacimiento”, “Pagador (Desc)”, “IEMA inlier”, “Estancias evitables” por ser columnas no informativas y con valores faltantes (NaN) no relevantes para la modelación.

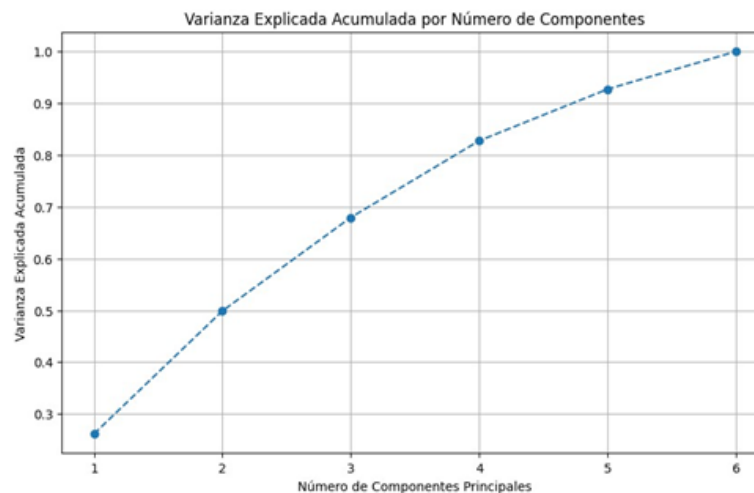
A la variable categórica “Pagador (Cod)” se le aplicó una técnica “One Hot Encoder”, que nos permite convertirla a un formato categórico binario y pueda ser interpretada por los modelos, el efecto de este proceso generó un aumento significativo del número de variables pasando de 16 columnas a 98.

El nombre de la variable “Motivo de Egreso descripción” se cambió por “Estado al Alta”, igualmente estaba desbalanceada con 2,978 Fallecidos y 61,506 Vivos, hallazgo que se abordó con una técnica de balanceo SMOTE (Synthetic Minority Over-sampling Technique) generando ejemplos en la clase minoritaria con los siguientes parámetros “sampling_strategy='auto'” que por defecto equilibra la clase minoritaria con la mayoritaria y se dejó por defecto “k_neighbors=5”, teniendo presente que este procesado solo aplica para los datos usados para entrenamiento.

El escalado estandarizado se aplicó a las variables predictoras usando “StandardScaler”, asegurando media cero y desviación estándar uno.

Por el efecto ya descrito de la binarización de la variable pagador (cod), que generó 82 variables adicionales y el tamaño de la base de datos, se aplicó una técnica de reducción de dimensionalidad, “Análisis de Componentes Principales” (PCA) conservando el 95% de la varianza explicada, lo cual permite mejorar la eficiencia computacional y reducir la complejidad a los modelos, usando los primeros cinco componentes que explican más del 90% de la varianza (ver gráfico 2).

Gráfico 2, Análisis PCA.



Para la modelación se formularon cuatro tareas de clasificación binaria:

1. Estancia hospitalaria prolongada: estancia mayor al percentil 3.
2. Costos hospitalarios altos: definido como costos superiores a la mediana.
3. Egresos mensuales elevados: meses con un número de egresos superior a la mediana mensual.
4. Estado al alta (sobrevida): Vivo = 0, Muerto = 1 desde la variable *Motivo de Egreso (Descripción)*.

Para el entrenamiento y validación de los modelos se aplicó validación cruzada con 5 folds, preservando la proporción de clases como previamente se había indicado.

6.1 DESARROLLO DE MODELOS Y AJUSTE DE HÍPER-PARÁMETROS

Con el objetivo de predecir los distintos desenlaces hospitalarios planteados, se implementaron modelos de clasificación supervisados, utilizando algoritmos de **Regresión Logística**, **Random Forest**, Máquinas de Vectores de Soporte (**SVM**), método de ensamble **XGBoost**, y una **Red Neuronal**.

Se definieron las siguientes métricas de desempeño para la comparación y selección del modelo:

- Exactitud
- Matriz de confusión
- F1 – Score

- Curva ROC – AUC
- Curva de precisión – Recall para los modelos de SVM y red neuronal.

Con el fin de comprender la contribución individual de cada variable al resultado de los modelos, se usó “SHAP” (SHapley Additive exPlanations), una técnica de interpretabilidad basada en teoría de juegos, que permite explicar la salida de cualquier modelo de aprendizaje automático. SHAP asigna a cada característica un valor de contribución (valor de Shapley) y cuantifica su impacto en la predicción del modelo. Esta metodología no solo proporciona interpretabilidad global del modelo (identificando las variables más importantes), sino también que permite entender la predicción para cada observación específica. Para identificar las variables más influyentes en la predicción se generaron gráficos “Beeswarm” que evalúan la dirección del efecto y la variabilidad del impacto de cada variable entre los diferentes modelos.

Es importante resaltar que, tanto en el ámbito clínico como en el administrativo, la interpretabilidad del modelo es un aspecto relevante en términos de credibilidad y usabilidad futura, por lo que se tendrá en cuenta al momento de evaluar los resultados, el concepto de *Mimic learning* (Aprendizaje por Imitación) que asume los modelos Shadow como los más interpretables (9).

6.1.2 MODELO DE REGRESIÓN LOGÍSTICA

- Penalización ElasticNet (solver= ‘saga’, l1 ratio=0.5).
- Para la predicción de egresos mes, se utilizó un modelo de Regresión de Poisson, que nos permite predecir conteos.

6.1.3 Modelo Random Forest

- n_estimadores = 100
- random_state=42 como clasificador robusto no lineal.

6.1.4 MODELO MAQUINAS DE VECTORES DE SOPORTE (SVM) KERNEL LINEAL.

- Modelo SVM (kernel='lineal')
- SVC (probability=True).

6.1.5 MODELO DE ENSAMBLE, EXTREME GRADIENT BOOSTING (XGBOOST).

- random_state=42,
- n_jobs=-1,
- use_label_encoder=False,
- eval_metric='logloss'

6.1.6 MODELO DE RED NEURONAL.

Se entrenó una red neuronal feedforward completamente conectada, utilizando la biblioteca Keras con backend en TensorFlow. La arquitectura fue diseñada específicamente para tareas de clasificación binaria y estuvo compuesta por las siguientes capas:

- Una **capa de entrada densa** con 90 neuronas y función de activación ReLU, que recibe el número de variables equivalente al número de componentes principales resultantes del PCA.
- Una **capa Dropout** con una tasa de abandono del 30% (Dropout=0.3) para mitigar el sobreajuste.
- Una segunda **capa oculta densa** con 32 neuronas y activación ReLU, seguida por otra capa Dropout=0.3.
- Una **capa de salida** con una sola neurona y función de activación **sigmoide**, adecuada para clasificación binaria.
- **Función de pérdida:** binary_crossentropy
- **Optimizador:** Adam con tasa de aprendizaje ajustada a 0.001

El entrenamiento se ejecutó con un máximo de 50 épocas, utilizando un esquema de validación cruzada estratificada de 5 folds, e incorporando EarlyStopping con patience = 5 y restauración de los mejores pesos, basado en la métrica de validación.

6.1.7 MODELO SARIMA.

La predicción de “Egresos Mes” es una serie temporal, por lo que implementamos un modelo SARIMA (AutoRegressive Integrated Moving Average) que permite predecir series de tiempo, para esto realizamos una prueba de estacionariedad (Dickey-Fuller) con un valor de p-value: 0.00088, lo cual confirma que la serie es estacionaria.

los parámetros (p,d,q)(P,D,Q) del modelo, se definieron usando la librería pmdarima. La función auto_arima busca automáticamente la mejor combinación de parámetros que minimiza el Criterio de Información de Akaike (AIC), simplificando el proceso. Para el entrenamiento del Modelo, Dividimos los datos cronológicamente en un conjunto de entrenamiento que corresponden a los primeros 24 meses y otro de prueba con los 12 meses restantes, el modelo final seleccionado fue un SARIMA (1, 1, 1)(1, 0, 1) 12. Que corresponde a la siguiente descripción: El pronóstico de un mes se basa en el valor del mes anterior (**p=1**), la serie fue diferenciada una vez para eliminar la tendencia y hacerla estacionaria (**d=1**) y el modelo se ajusta utilizando el error de predicción del mes anterior (**q=1**). A nivel estacional, el pronóstico toma en cuenta el número de egresos del mismo mes del año anterior (**P=1**). No fue necesaria una diferenciación estacional, lo que sugiere que el ciclo anual era estable (**D=0**) y se ajusta utilizando el error de predicción del mismo mes del año anterior (**Q=1**).

6.2 DEFINICIÓN DE HIPER PARÁMETROS

Cada modelo se complementó con un código “RandomizedSearchCV” para la búsqueda de los mejores hiperparámetros buscando optimizar las métricas planteadas, se describen a continuación (11).

REGRESIÓN LOGÍSTICA:

- 'model__penalty': ['l1', 'l2']
- 'model__C': np.logspace(-4, 4, 100)
- 'model__solver': ['liblinear']
- n_iter=20
- cv=5
- random_state=42
- verbose=1

RANDOM FOREST:

- 'model__n_estimators': [100, 200, 300, 500]
- 'model__max_depth': [10, 20, 30, None]
- 'model__min_samples_leaf': [1, 2, 4, 6]
- 'model__max_features': ['sqrt', 'log2', None]

XGBOOST:

- 'model__n_estimators': [100, 200, 300, 500],
- 'model__learning_rate': [0.01, 0.05, 0.1, 0.2]
- 'model__max_depth': [3, 5, 7, 10]
- 'model__subsample': [0.7, 0.8, 0.9, 1.0]

MÁQUINAS DE VECTORES DE SOPORTE (SVM, KERNER LINEAR):

- 'model__C': np.logspace(-2, 3, 10)

RED NEURONAL:

Para la búsqueda de hiper-parámetros se usó *Keras Tuner* con el algoritmo de *Random Search* para optimizar una red neuronal densa. La función de construcción del modelo permitió ajustar de forma automática el número de neuronas en las capas ocultas, las tasas de *dropout* y la tasa de aprendizaje del optimizador Adam.

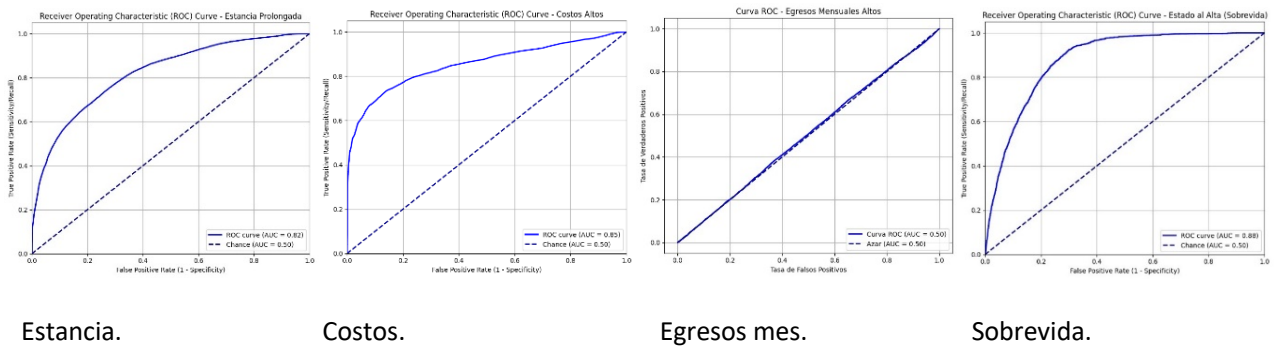
La métrica objetivo “*val_accuracy*” y “*EarlyStopping*” permiten prevenir el sobreajuste y los mejores hiper-parámetros obtenidos.

7.0 RESULTADOS

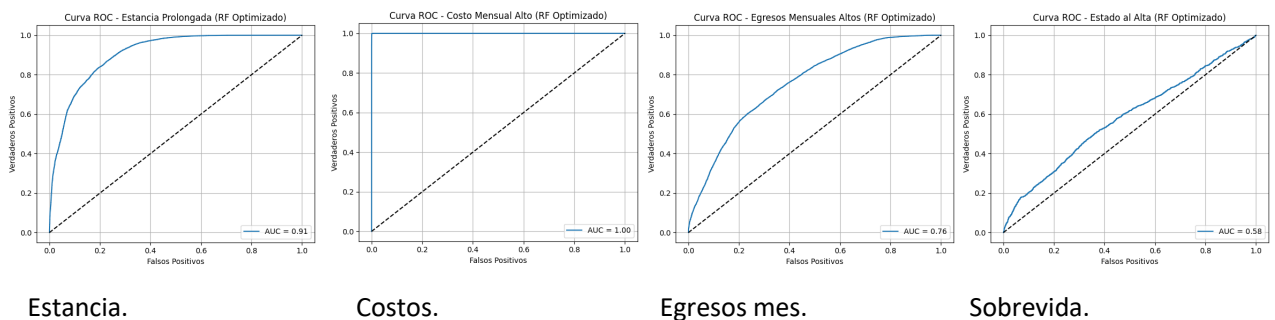
Los resultados se presentan comparando las métricas definidas para los diferentes modelos y gráficamente con la curva de AUC para cada una de las variables objetivo. Seguidamente se hace un comparativo entre las variables de mayor peso según modelo y desempeño, usando u técnica de interpretabilidad SHAP.

Las gráficas se presentan en orden de acuerdo con las variables objetivo: Estancia, Costos, Egresos mes y Estado al alta, respectivamente.

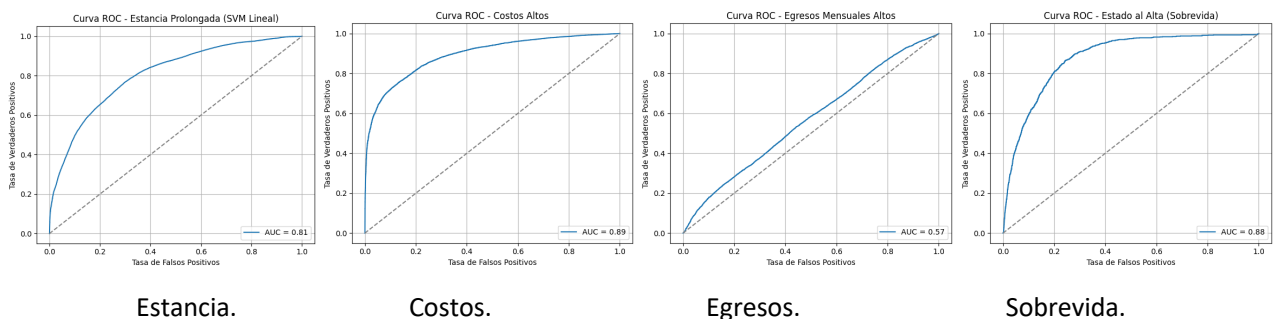
7.1 REGRESIÓN LOGÍSTICA



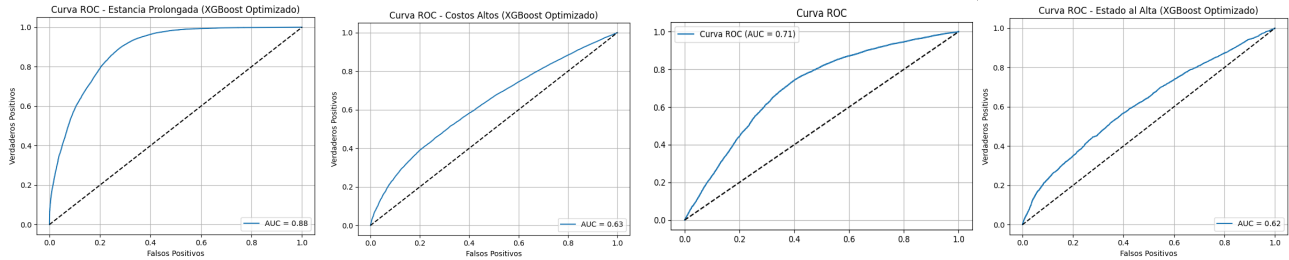
7.2 RANDOM FOREST



7.3 MÁQUINAS DE VECTORES DE SOPORTE (SVM) KERNEL LINEAR.



7.4 Modelo XGboost



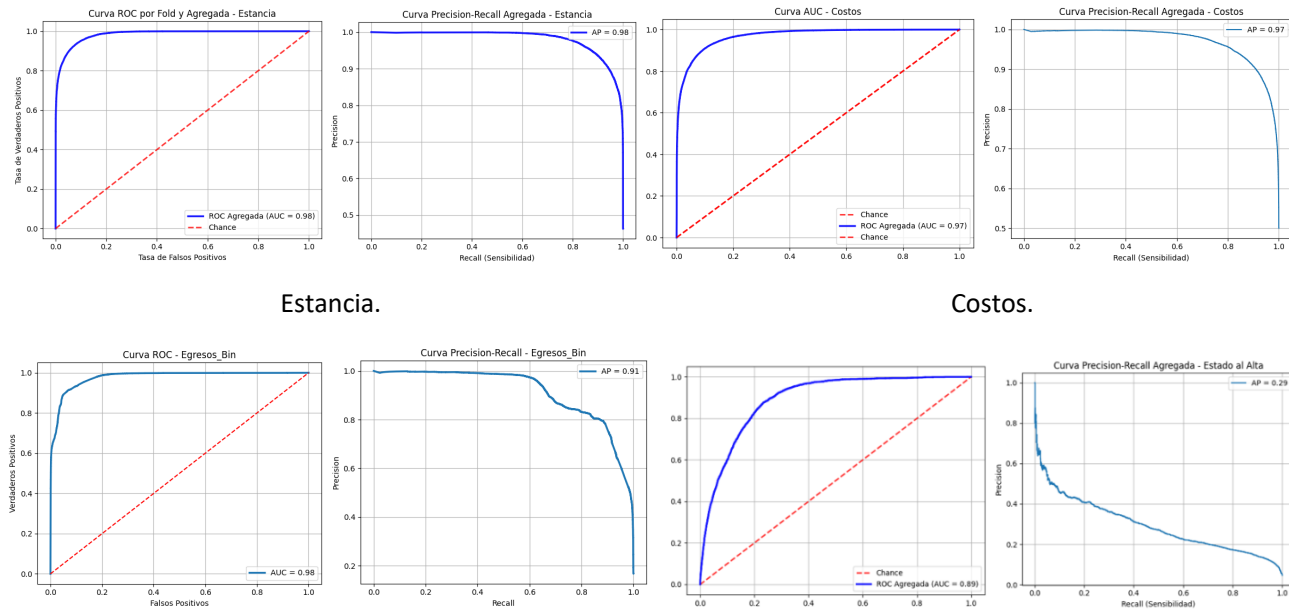
Estancia.

Costos.

Egresos mes.

Sobrevida.

7.5 Red Neuronal (Curva AUC y Precision – Recall).



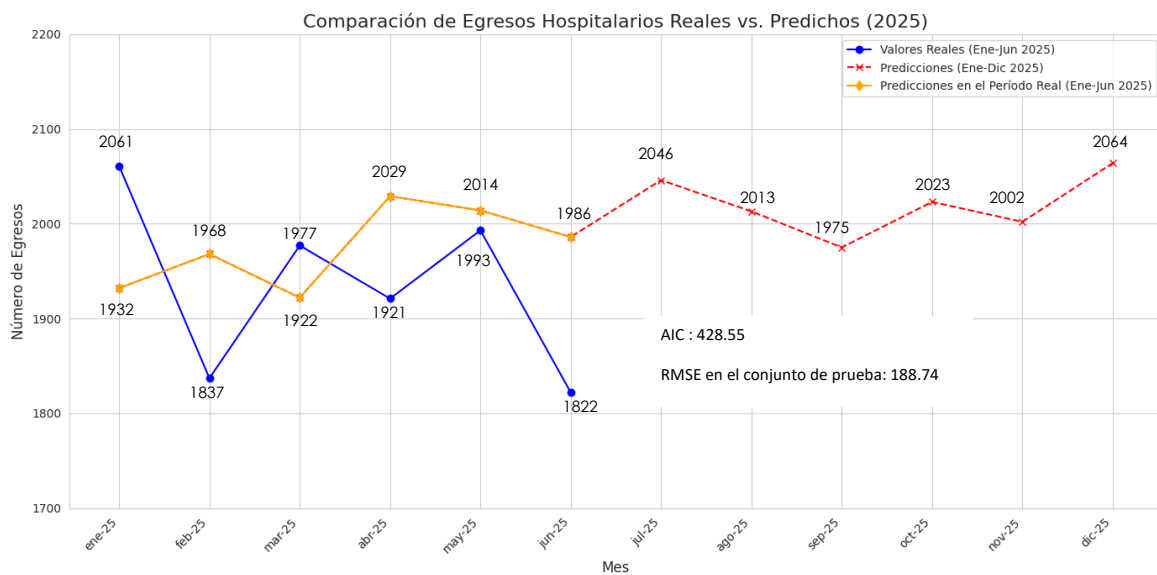
Estancia.

Costos.

Egresos mes.

Sobrevida.

7.6 Modelo SARIMA



Predicción egresos mes 2025.

7.7 TABLAS DE RESULTADOS

Tabla 3. Estancia mayor a la media.

Modelos -Estancia	Exactitud	Precisión	F1 Score	AUC
Regresión logística	0.7388	0.74	0.7367	0.8213
Random-Forest	0.8135	0.82	0.8187	0.9071
SVM	0.7323	0.73	0.7034	0.8083
XGBoost	0.8018	0.81	0.8057	0.8833
Red Neuronal	0.9264	0.93	0.9211	0.9835

Tabla 4. Costos, mayor a la media:

Modelos - Costos	Exactitud	Precisión	F1 Score	AUC
Regresión logística	0.7968	0.80	0.7959	0.8536
Random-Forest	0.9951	0.99	0.9951	0.9997
SVM	0.8091	0.80	0.8048	0.8936
XGBoost	0.5940	0.60	0.5659	0.6300
Red Neuronal	0.9046	0.90	0.9051	0.9646

Tabla 5. Egresos, mayor a la media mensual:

Modelos - Egresos	Exactitud	Precisión	F1 Score	AUC
Regresión logística	0.4820	0.50	0.4304	0.5045
Random-Forest	0.6835	0.68	0.6488	0.7572
SVM	0.5434	0.54	0.5059	0.5668
XGBoost	0.6645	0.67	0.6409	0.7081
Red Neuronal	0.9255	0.85	0.8051	0.9775

Tabla 6. Estado al alta, vivo – muerto:

Modelos- Estado al alta	Exactitud	Precisión	F1 Score	AUC
Regresión logística	0.7353	0.56	0.5370	0.8787
Random-Forest	0.7313	0.51	0.1104	0.5825
SVM	0.7611	0.57	0.2442	0.8784
XGBoost	0.6906	0.52	0.1201	0.6154
Red Neuronal	0.7696	0.58	0.2779	0.8890

TABLA 7. MODELO SARIMA PARA PREDICCIÓN DE EGRESOS MES

	Coef	Std err	P> z
Intercept	9.9655	5,8640	0,0890
ma.L1	-0.7364	0,1340	0,0000
ma.S.L12	0.4498	0,4300	0,2950
sigma2	7462.221	2356.317	0,0020

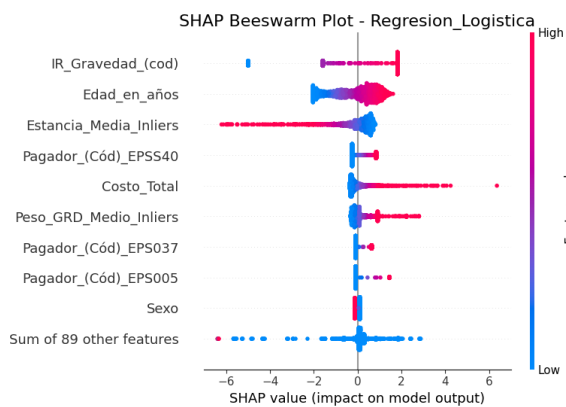
Las métricas para este modelo son diferentes y no permiten hacer una comparación con los anteriores modelos, por lo cual se describe a continuación, la interpretación de estas: AIC del modelo SARIMA: 428.55 lo que sugiere poco ajuste y el RMSE en el conjunto de prueba: 188.74 se puede interpretar como baja precisión en las predicciones.

- **intercept (Valor: 9.9655, P-valor: 0.089):** El intercepto no es estadísticamente significativo. Esto significa que el modelo podría funcionar igual de bien sin el análisis de estacionalidad.
- **ma.L1 (Valor: -0.7364, P-valor: 0.000)** Este es el coeficiente de Media Móvil (MA) no estacional de orden 1. Su p-valor es 0.000, lo cual es altamente significativo y significa que el error de la predicción del mes anterior tiene una influencia directa y significativa en la predicción del mes actual.
- **ma.S.L12 (Valor: 0.4498, P-valor: 0.295)** el coeficiente de Media Móvil Estacional (MA). Indica que es estacional y el L12 que su ciclo es de 12 meses. Su p-valor sugiere que este componente estacional no es estadísticamente significativo. Aunque auto_arima lo incluyó (probablemente porque mejoraba ligeramente el AIC), los datos no muestran una evidencia fuerte de que el error del mismo mes del año anterior sea un buen predictor.
- **sigma2 (Valor: 7462.2211, P-valor: 0.002), la varianza de los residuos** (los errores del modelo). El p-valor cercano a cero simplemente confirma que la varianza existe, lo cual es normal.

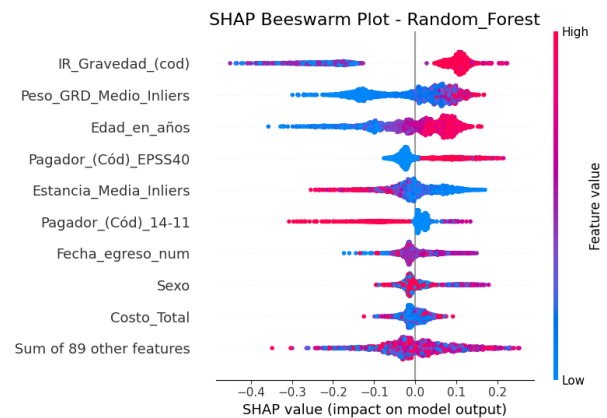
7.7 COMPARATIVO SHAP DE LOS MODELOS

Para efectos de eficiencia computacional se redactó un código usando 6000 muestras de la base de datos, 10 épocas y limitamos la muestra de validación a 1000. Los resultados de este comparativo se describen a continuación con graficas representativas de las variables con mayor peso, se detallan en la tabla 7.

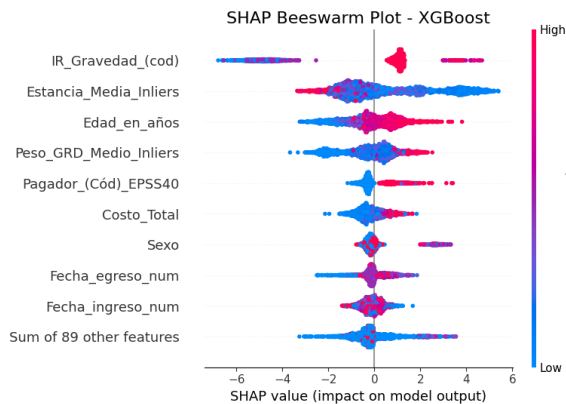
Regresión logística.



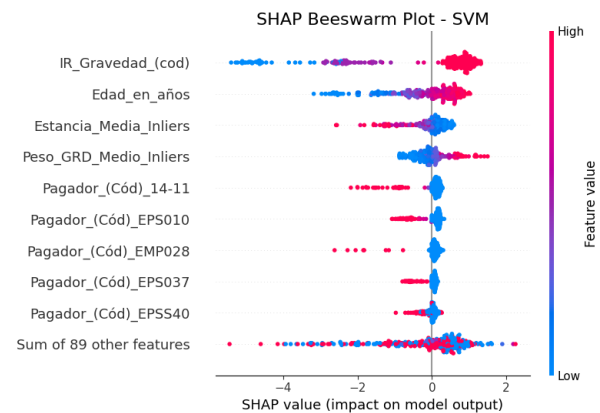
Random Forest.



XGBoost.



SVM.



Red Neuronal.

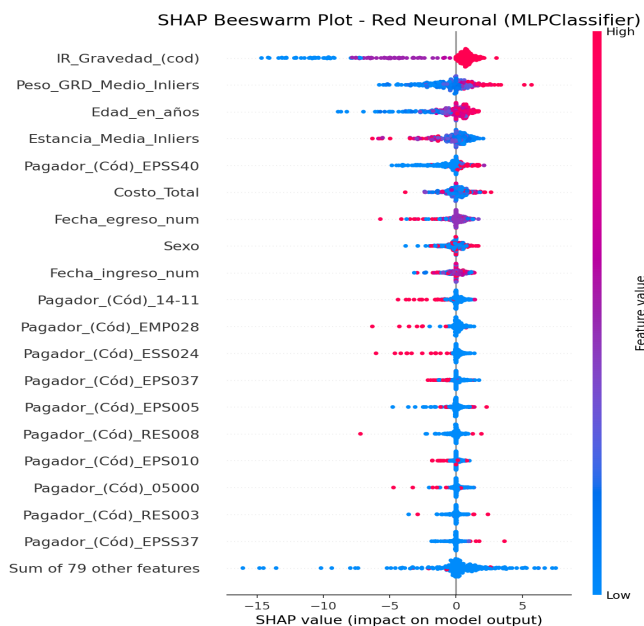


Tabla 7. Análisis SHAP según modelo y variable.

Modelo	Top Variables	Tipo de relación	Interpretabilidad	Capacidad predictiva
Red Neuronal	IR Gravedad, Peso GRD, Edad.	No Lineal	Baja	Alta
Random Forest	IR Gravedad, Edad, Peso GRD	No Lineal	Media	Alta
Regresión Logística	IR Gravedad, Edad, Estancia.	Lineal	Muy Alta	Media
XGBoost	IR Gravedad, Estancia, Edad.	No Lineal	Media	Muy Alta
SVM	IR Gravedad, Edad, Estancia.	Lineal	Alta	Media

La interpretación SHAP por modelo muestra como para la Red Neuronal, las variables más influyentes encontradas fueron: IR Gravedad, Peso GRD, Edad, Estancia, Pagador EPS40, el modelo captura relaciones complejas y no lineales entre variables y predicción con alto grado de dispersión, y el más bajo desempeño lo encontramos en la moderada capacidad para diferenciar entre Vivos vs. Muertos.

Para la Regresión Logística, las variables más influyentes fueron, IR Gravedad, Edad, Estancia. Encontramos que es el modelo con el menor número de variables con gran impacto, lo que lo hace más sencillo y explicativo, aunque pierde precisión en relaciones complejas. Random Forest tiene como variables más importantes IR Gravedad, Edad, Peso GRD, con buena capacidad explicativa y manejo de relaciones no lineales.

En XGBoost, las variables IR Gravedad, Estancia, Edad, Peso GRD y Pagador EPS40, fueron las de mayor impacto, captura interacciones no lineales y muestra patrones diferenciados por el nivel de valor de cada variable, podría recomendarse para modelos tabulares por su precisión y capacidad interpretativa.

Los predictores dominantes para las Maquinas de vectores de soporte, fueron: FI Gravedad, Edad, Estancia, Peso GRD. Con menor dispersión y un modelo más lineal y estable en su interpretación, útil para interpretaciones directas, pero menos potente en relaciones complejas como lo vemos en la predicción de egresos y estado al alta.

8.0 DISCUSIÓN

Este estudio exploratorio, evaluó el desempeño de seis modelos de aprendizaje automático para la predicción de variables claves en la gestión hospitalaria, los resultados revelaron un rendimiento heterogéneo entre los modelos y las tareas propuestas, ofreciendo información valiosa sobre la capacidad predictiva y su interpretabilidad (Gráfico 8).

Con respecto a la **predicción de la estancia**, encontramos que la red neuronal obtuvo el mejor desempeño, con una exactitud del 92.6%, y alcanzando un AUC de 0.9835, lo que sugiere una alta capacidad discriminatoria entre interacciones complejas no lineales de variables como, edad, gravedad y tipo de ingreso.

El modelo Random Forest también mostró un excelente balance, con un AUC de 0.9071, y a su vez el modelo explicativo SHAP indica una buena capacidad para identificar variables clave como gravedad, edad y diagnóstico, lo que lo posiciona como un modelo robusto y relativamente interpretable. XGBoost logró un AUC de 0.8833, confirmando su potencia en precisión, aunque con interpretabilidad intermedia. Por otro lado, modelos como Regresión Logística y SVM (kernel lineal) presentaron menor rendimiento $AUC < 0.85$, pero alta claridad en sus decisiones, y los hace valiosos en contextos donde la transparencia del modelo es crítica.

Los resultados según la variable objetivo “Estancia prolongada” vemos como los múltiples diagnósticos activos, IR Gravedad, tienen alta probabilidad de nuevas estancias prolongadas. La edad emerge como factor clave y especialmente en el modelo de Random-Forest, donde los pacientes de mayor edad tienden a permanecer más tiempo hospitalizados.

Para la predicción de **Costos hospitalarios**, los resultados muestran que el modelo Random Forest obtuvo el mejor desempeño, con una exactitud del 99.5% y un AUC de 0.9997, lo que sugiere un ajuste casi perfecto sobre los datos de entrenamiento. Esto puede indicar una excelente capacidad predictiva, sin descartar la posibilidad de sobreajuste.

Por su parte, el modelo de Red Neuronal se ubicó como el segundo mejor en rendimiento, con una AUC de 0.9646, balanceando buena capacidad predictiva con menor riesgo de sobreajuste que el Random Forest. Este resultado, junto con su comportamiento en el análisis SHAP, lo posiciona como un modelo potente para este tipo de predicciones. La Regresión Logística mostró un desempeño aceptable $AUC = 0.8536$, y precisión de 0.80%, permite interpretabilidad, lo que puede ser útil en contextos donde se requiere justificar las decisiones basadas en costos. El modelo SVM obtuvo resultados similares a la regresión logística $AUC = 0.8936$, mientras que XGBoost mostró un rendimiento significativamente más bajo AUC de 0.63, lo que sugiere que estos dos algoritmos podrían no ser los más adecuados para la predicción de costos en este contexto específico, posiblemente debido a la naturaleza lineal o ruido en las variables predictoras.

El “IR GRD”, la “estancia media”, Egreso Medico, fueron predictores robustos en todos los modelos para la estimación de los costos. Igualmente la “edad avanzada” y “mayor gravedad clínica” se asocian a mayores costos hospitalarios indicando mayor comorbilidad y tratamientos complejos.

Respecto a la **predicción de los Egresos hospitalarios**, los modelos presentaron un comportamiento heterogéneo. La Red Neuronal nuevamente lideró en desempeño con una exactitud del 92.6%, AUC de 0.9775. La Regresión Logística mostró una baja capacidad explicativa con un AUC de 0.5045, y una precisión de 0.50, lo que la convierte en una alternativa poco razonable para este tipo de tarea, en contraste, XGBoost y Random Forest presentaron rendimientos más discretos AUC entre 0.70 y 0.75, y podría estar relacionado con una mayor dificultad de los modelos para capturar patrones relevantes en los datos. Finalmente, las SVM tuvieron pobre desempeño con $AUC = 0.5668$,

lo que indica que este modelo no logra capturar adecuadamente la estructura subyacente del problema, entre las explicaciones posibles esta la temporalidad inherente en las características predictoras.

La estimación de los Egresos mensuales y el Estado al Alta, fueron particularmente los más desafiantes, posiblemente debido a la influencia de factores estacionales no modelados explícitamente y la composición de la base de datos, con información predominantemente de datos demográficos y administrativos. Sin embargo la implementación de un modelo SARIMA permitió hacer las predicciones, con unas métricas que demuestran la capacidad y potencia de este modelo para series temporales.

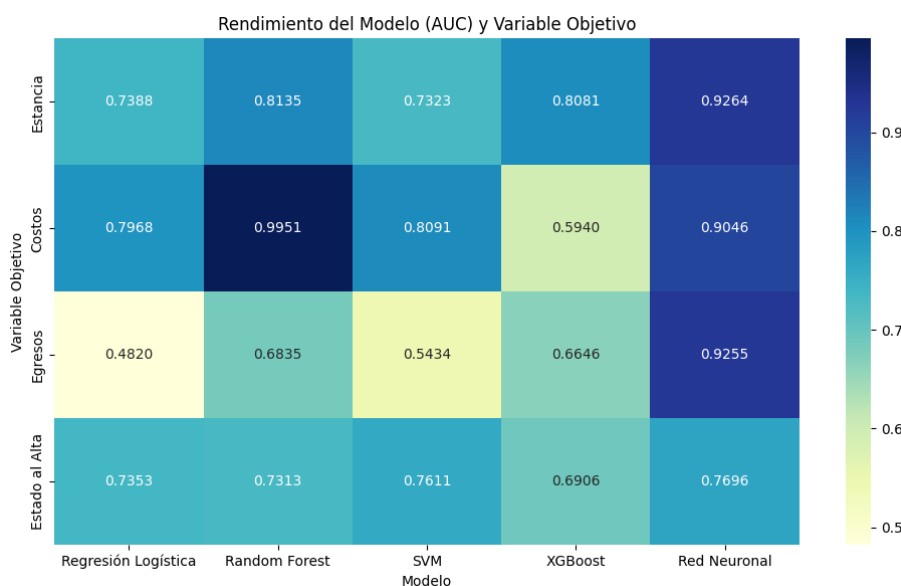
Para la **predicción de Estado al Alta**. La tarea de predecir la sobrevida hospitalaria mostró ser especialmente compleja para todos los modelos. Aunque las métricas de exactitud oscilan entre 69% y 77%, los valores de F1 Score y precisión son significativamente bajos, lo cual refleja grandes dificultades para predecir correctamente la clase minoritaria (pacientes que fallecen). La Red Neuronal alcanzó el mejor desempeño general AUC = 0.8890, seguida de la Regresión Logística con AUC = 0.8787, y en tercer lugar las SVM con un AUC = 0.8784 y todas presentaron F1 Scores bajos (< 0.53), indicando una capacidad muy limitada para identificar verdaderos positivos sin generar una gran cantidad de falsos positivos. A pesar de que la red neuronal logra un valor de AUC significativo, su precisión en la clase minoritaria AP = 0.29 es baja, lo que limita su uso clínico directo sin medidas de compensación adicionales (ej: ajuste de umbral o calibración).

El Random Forest y XGBoost mostraron muy bajo desempeño, con AUC menores a 0.62, lo cual indica que no son adecuados para esta tarea en su forma actual, probablemente por el fuerte desbalance de clases o por la escasa diferenciación de patrones en los datos disponibles.

Este conjunto de resultados sugiere que la sobrevida podría requerir un reentrenamiento específico, el manejo del desbalance de clases (SMOTE, ajuste de pesos, calibración) o incluso la incorporación de variables clínicas más relevantes y con mejor capacidad explicativa.

La edad, el tipo de IR_GRD, el nivel de gravedad al ingreso y los días de estancia hospitalaria, mostraron un efecto positivo sobre la probabilidad del desenlace, lo que valida tanto el razonamiento clínico como la estabilidad del modelo.

Gráfico 8. Mapa de Calor Modelo vs Variable objetivo.



La relación interpretabilidad versus precisión es un elemento clave para la adopción de modelos de aprendizaje automático en el ámbito administrativo. El modelo de regresión logística si bien fue ligeramente menos preciso, es

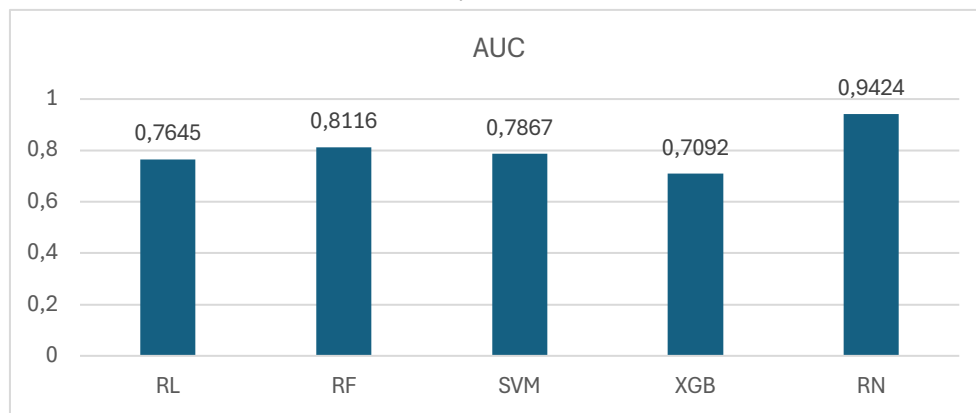
particularmente útil para la interpretación de las variables y como referencia lineal frente a otros modelos de mayor complejidad. Ahora bien, los modelos Random Forest y la Red neuronal, permitieron explorar la importancia relativa de las variables predictoras y obtener mayor desempeño en relaciones no lineales. Los resultados muestran que los enfoques de modelado son efectivos para predecir desenlaces hospitalarios clave, y se resumen en la tabla 8.

Tabla 8. Resumen comparativo (Variable – Modelo – Métricas).

Variable	Modelo con mejor AUC	AUC	Modelo con mejor F1 score	F1 Score	Modelo con mejor Exactitud	Exactitud
Estancia	Red Neuronal	0.9264	Red Neuronal	0.9211	Red Neuronal	0.9264
Costos	Random Forest	0.9951	Random Forest	0.9951	Random Forest	0.9951
Egresos Mes	Red Neuronal	0.9255	Red Neuronal	0.8051	Red Neuronal	0.9255
Estado al Alta	Red Neuronal	0.7696	Red Neuronal	0.2779	Red Neuronal	0.7696

Los modelos, Random Forest y las Redes Neuronales superaron consistentemente a los demás modelos, es de anotar que el modelo de red neuronal tuvo un bajo desempeño para la predicción del “Estado al alta”, también se destaca la capacidad del modelo SARIMA, para la predicción de los egresos mes, que corresponde a una serie temporal, superando todos los demás modelos evaluados para la predicción de esta variable. En el grafico 9, vemos el promedio de la AUC para cada uno de los modelos.

Gráfico 9.
AUC promedio.



Es importante destacar como el uso de PCA y SMOTE, fueron fundamentales para mejorar la estabilidad de los modelos, especialmente ante datos desbalanceados, además se recomienda la inclusión de técnicas explicativas como SHAP para mejor interpretabilidad del modelo y la relación entre modelo y variables.

Los resultados de este proyecto, respaldan el papel de los modelos basados en aprendizaje automático, como herramientas útiles para la estimación de necesidades y con ello el manejo óptimo de los recursos hospitalarios. Publicaciones recientes en este ámbito van en la misma línea, algunos con una metodología similar, reportan datos comparables con los nuestros, en todos estos se destaca la importancia de la calidad de la información disponible, para obtener buenos resultados(4), (5), (6).

Las principales limitaciones encontradas en este proyecto estuvieron en la dificultad para la estimación de la Sobrevida y la predicción de Egresos Mes, fue compleja, y como regla general todos los modelos tuvieron bajo rendimiento. Entre las posibles explicaciones tenemos la naturaleza cambiante de estas variables, además de estar sujetas a factores temporales, muy posiblemente nuestra base de datos no tiene las suficientes variables clínicas informativas para hacer una estimación más exacta. También vimos como la carga computacional de algunos modelos como las máquinas de Vector de Soporte, limitaron notoriamente su uso en especial para el modelamiento de los egresos y estado al alta.

9.0 CONCLUSIONES

Los resultados de este estudio, demuestran como los modelos de aprendizaje automático, entrenados con datos demográficos y administrativos de un entorno hospitalario, pueden estimar con precisión variables claves para la gestión de recurso, validando la hipótesis del proyecto. Las métricas subrayan que no existe un único modelo para todas las tareas; la elección óptima depende de un balance entre la capacidad predictiva y la necesidad de interpretabilidad.

10.0 BIBLIOGRAFÍA

- [1] Arjun Panesar. Machine learning an AI for healthcare. Second edition, 2021. Apress.
- [2] Ibrahim Adedeji Adeniran, Christianah Pelumi Efunniyi, Olajide Soji Osundare, and Angela Omozele Abhulimen. Data-driven decision-making in healthcare: Improving patient outcomes through predictive modeling; International Journal of Scholarly Research in Multidisciplinary Studies, 2024, 05(01), 059–067. <https://doi.org/10.56781/ijrms.2024.5.1.0040>
- [3] Milan Toma , Ong Chi Wei. Predictive Modeling in Medicine. Encyclopedia 2023, 3, 590–601. <https://doi.org/10.3390/encyclopedia3020042>.
- [4] Pfof Andr., Sheng-Chieh Lu and Chris Sidey-Gibbons. Machine learning in medicine: a practical introduction to techniques for data pre-processing, hyperparameter tuning, and model comparison: BMC Medical Research Methodology (2022) 22:282, <https://doi.org/10.1186/s12874-022-01758-8>
- [5] Yazan Alnsour, Marina Johnson, Abdullah Albizri, Antoine Harfouch. Predicting Patient Length of Stay Using Artificial Intelligence to Assist Healthcare Professionals in Resource Planning and Scheduling Decisions. Journal of Global Information Management Volume 31: Issue 1. DOI: 10.4018/JGIM.323059
- [6] Syed Hyder Badshah, AI-Driven Decision-Making in Healthcare Administration: Exploring Predictive Models for Resource Allocation. Journal of Artificial Intelligence and Health; London Edition. 2024 Vol. 1, No. 1, pp- 21-27. <https://journalaihealth.uk>
- [7] Sandra Eloranta, Magnus Boman. Predictive models for clinical decision making: Deep dives in practical machine learning; Journal of Internal Medicine, 2022, 292; 278–295. DOI: [10.1111/joim.13483](https://doi.org/10.1111/joim.13483)
- [8] Mohammed Badawy, Nagy Ramadan and Hesham Ahmed Hefny. Healthcare predictive analytics using machine learning and deep learning techniques: a survey. Journal of Electrical Systems and Inf Technol (2023) 10:40 <https://doi.org/10.1186/s43067-023-00108-y>
- [9] Farah Shamout, Tingting Zhu, and David Clifton. Machine Learning for Clinical Outcome Prediction; IEEE Reviews in biomedical engineering, vol 14. 2021. DOI: [10.1109/RBME.2020.3007816](https://doi.org/10.1109/RBME.2020.3007816)
- [10] Andre Pfof, Sheng-Chieh Lu, Chris Sidey-Gibbons. Machine learning in medicine: a practical introduction to techniques for data, pre-processing, hyperparameter tuning, and model comparison. BMC (Medical Research Methodology) 2022; 22: 282. <https://doi.org/10.1186/s12874-022-01758-8>
- [11] Amaya Mary Vincent, P. Jidesh. An improved hyperparameter optimization framework for AutoML system using evolutionary algorithms. Nature Scientific Reports 2023; 13:4737 <https://doi.org/10.1038/s41598-023-32027-3>
- [12] Samantha Cruz Rivera, Xiaoxuan Liu, An-Wen Chan, Alastair K Denniston, Melanie J Calvert, and The SPIRIT-AI and CONSORT-AI Working Group. Lancet Digital Health 2020; 2: e549 -560. [https://doi.org/10.1016/S2589-7500\(20\)30219-3](https://doi.org/10.1016/S2589-7500(20)30219-3)
- [13] Gary S Collins,1 Karel G M Moons,2 Paula Dhiman, et All. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods BMJ 2024;385:e078378. <https://doi.org/10.1136/bmj-2023-078378>
- [14] Shortliffe EH, Sepúlveda MJ. Clinical Decision Support in the Era of Artificial Intelligence. Jama. 2018;320 (21):2199-200.
- [15] Nadia Daneshvar, JD, MPH; Deepti Pandita, MD; Shari Erickson, MPH; Lois Snyder Sulmasy, JD; and, Matthew DeCamp, MD, PhD; for the ACP Medical Informatics Committee and the Ethics, Professionalism and Human Rights Committee. Artificial Intelligence in the Provision of Health Care: An American College of Physicians Policy Position Paper. Ann Intern Med 2024. doi:10.7326/M24-0146.