**Modelo predictivo de AKI postoperatoria en cirugía no cardíaca (UDAKI)**

Comprensión del Negocio

Nuestra propuesta se centra en desarrollar un modelo predictivo de AKI (Acute Kidney Injury) postoperatoria en pacientes programados para cirugía no cardíaca, utilizando técnicas de machine learning.

El objetivo principal es desarrollar una aplicación basada en el modelo generado con inteligencia artificial que, alimentado con datos de un paciente, genere como resultado su probabilidad de desarrollar lesión renal aguda postoperatoria (AKI).

Este modelo busca servir como una herramienta de apoyo en la toma de decisiones clínicas, permitiendo una identificación temprana de pacientes en riesgo y facilitando intervenciones preventivas.

Objetivos específicos

1. Desarrollar modelos de regresión logística, Random Forest, gradiente extremo (XG Boost) y Redes neuronales recurrentes para predecir la probabilidad de presentar AKI en los primeros 7 días del post operatorio.
2. Comparar el rendimiento de los modelos para predecir la aparición AKI post operatoria.
3. Construir una aplicación con una interfaz con el mejor modelo.

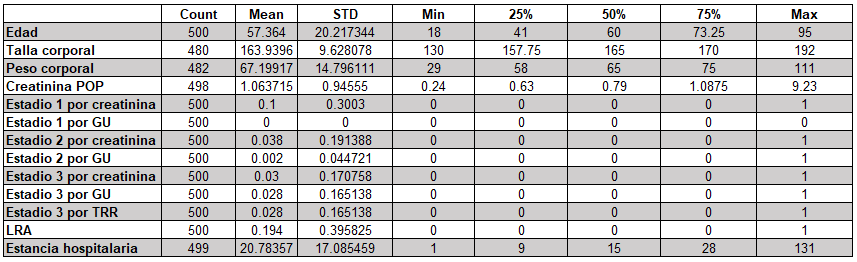
Comprensión de los Datos

El conjunto de datos para el desarrollo de los modelos se obtiene de un estudio de cohorte retrospectivo realizado por estudiantes de la especialización de anestesiología y reanimación de la universidad de Antioquia , en proceso de publicación, en pacientes mayores de 18 años atendidos entre del 1 de enero de 2020 al 31 de diciembre de 2022 en los hospitales Alma Máter de Antioquia y Pablo Tobón Uribe de la ciudad de Medellín, Colombia, que fueron llevados a cirugía no cardiaca y que tenían una medición de creatinina sérica basal hasta 30 días antes del procedimiento y un control dentro de 48 horas hasta 7 días postquirúrgicos.

La información fue obtenida mediante la revisión de las historias clínicas por parte de dos investigadores, a partir de bases de datos aportadas por ambas instituciones hospitalarias. La información se registró en la herramienta Redcap, posteriormente se convirtieron y unieron en una base de datos de Excel.

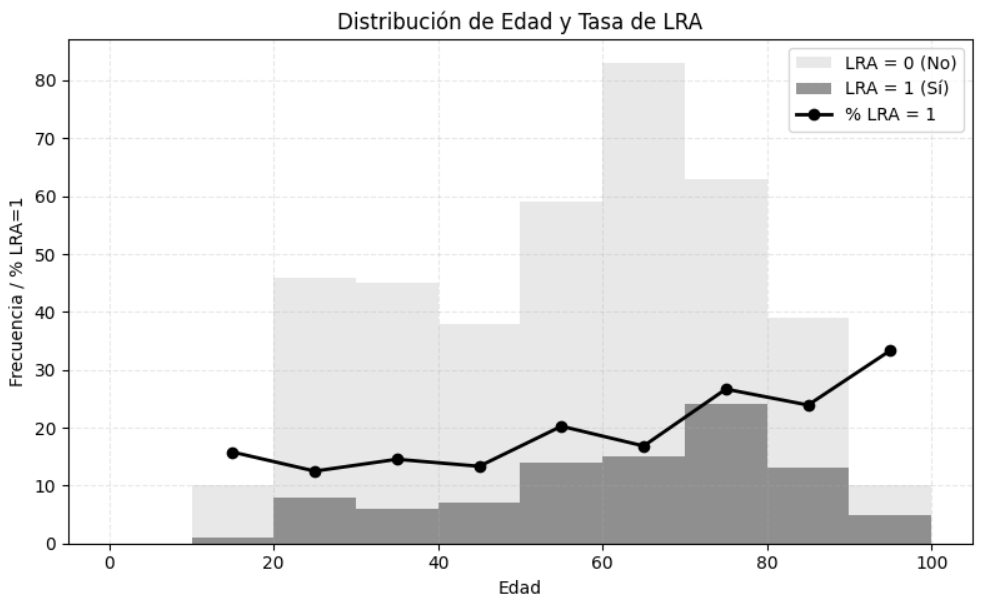
El conjunto de datos consta de 500 pacientes, las variables fueron: edad, sexo, índice de masa corporal, antecedentes patológicos como diabetes mellitus, hipertensión arterial, insuficiencia cardiaca crónica, neoplasia, infarto agudo al miocardio, cirrosis hepática, así como diagnósticos asociados como sepsis o choque séptico, rabdomiólisis, crisis hipertensiva, falla cardiaca descompensada además de uso de contraste endovenoso prequirúrgico, proteinuria preoperatoria, uso de antagonistas de los receptores de angiotensina II (ARA II) e inhibidores de la enzima convertidora de angiotensina (IECA), clasificación de american society of anesthesiologists (ASA), prioridad y tipo de la cirugía, técnica anestésica, ingreso a unidad de ciudadanos intensivos (UCI), sangrado intraoperatorio mayor o igual a 500 cc, transfusión de hemoderivados en el perioperatorio, uso de vasopresores en el intraoperatorio, lesión renal aguda post operatoria (LRA POP) según criterios KDIGO, estadios de la lesión renal estratificados en 1,2 y 3, días de estancia hospitalaria y muerte a los 28 días por cualquier causa o relacionada con LRA.

La media de la edad fue 57.4 años, de los cuales el 54% fueron hombres. Los antecedentes personales patológicos más comunes fueron hipertensión arterial (40%), seguido de neoplasia (maligna o del sistema nervioso central) (23%) y diabetes mellitus tipo 2 (15%).



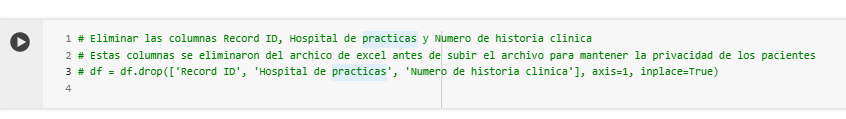
# La variable lesión renal aguda, es nuestra variable objetivo

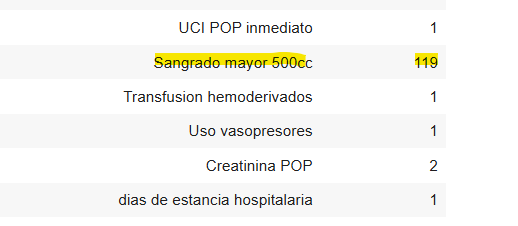
# tiene 94 casos positivos y casos positivos y 401 negativos, lo que de entrada plantea un conjunto de datos desbalanceado para la variable objetivo.

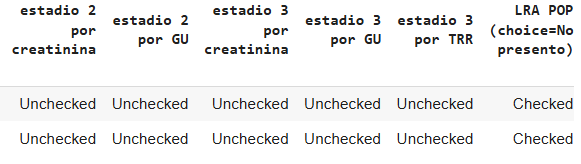


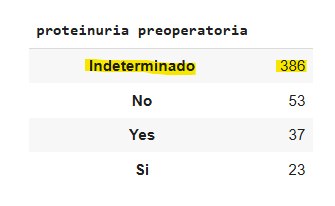
# Al revisar las variables se encontraron los siguientes problemas:

* El conjunto de datos tenía las variables Record ID, Hospital y Número de historia clínica del paciente, que pudieran facilitar la identificación de los pacientes
* La variable sangrado mayor a 500 cc faltan 119 datos, la imputación de esta variable es difícil porque el sangrado durante una cirugía está relacionado a factores del paciente como al tipo de cirugía. No se puede imputar por la moda, ni por vecinos cercanos.
* En las variables estadios, Unchecked significa NO ocurrencia del evento y Checked significa sí. En la variable LRA POP (choice=No presentó), Checked significa que no presento AKI, Unchecked significa que, si presento AKI, esta forma de registrar el valor dificulta la interpretación
* La variable proteinuria preoperatoria tiene 322 datos indeterminados



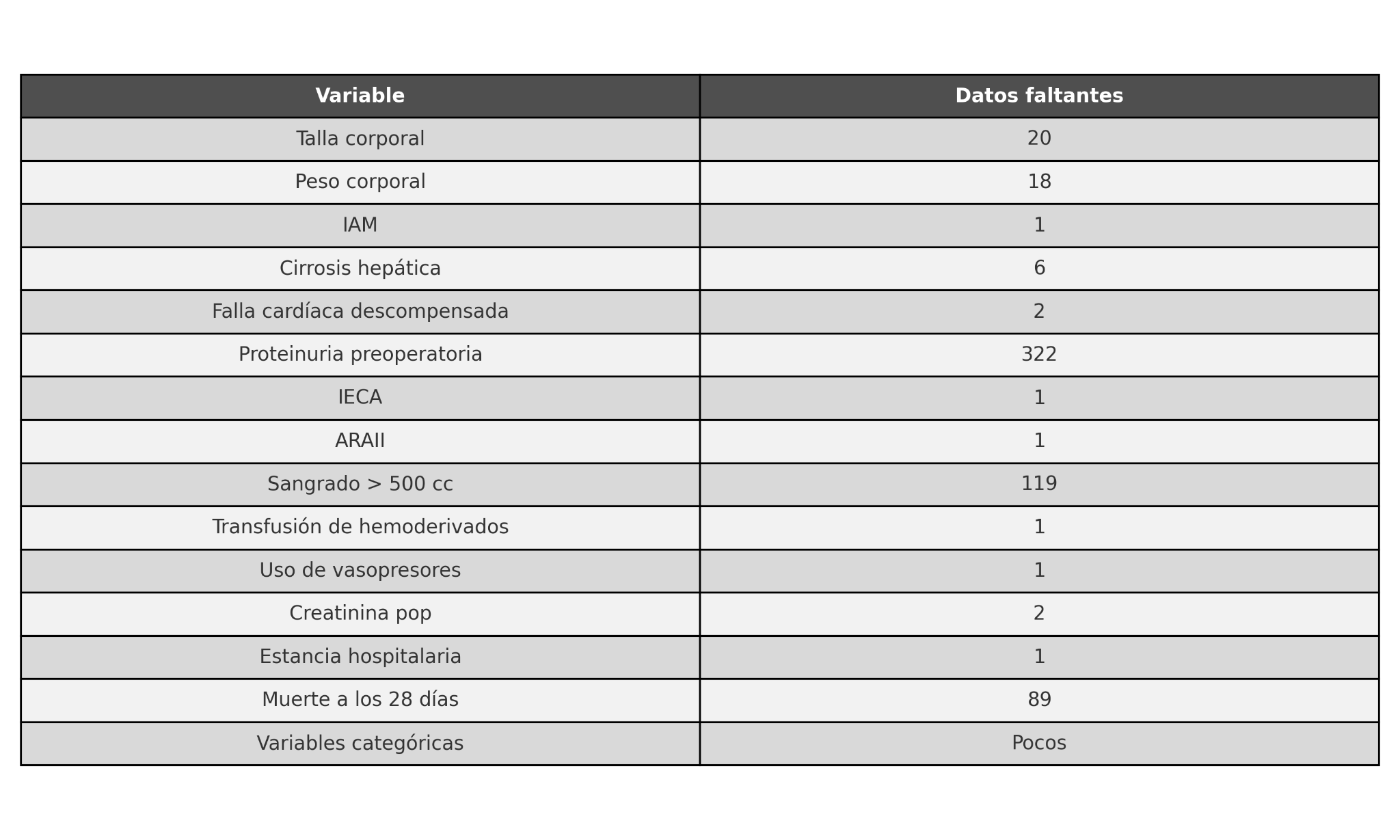






PROCESAMIENTO DE DATOS

**Datos faltantes:**



Manejo de los datos faltantes

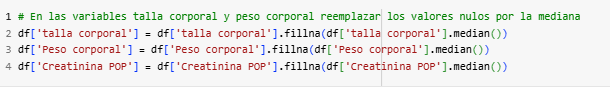
Muerte a los 28 días y relacionada con LRA: la variable mortalidad es una variable de resultado, dado que el objetivo del modelo es predecir la lesión renal aguda (LRA pop), y esta variable tiene gran número de datos faltantes, se decide eliminar.



La variable sangrado mayor a 500 cc faltan 119 datos, la imputación de esta variable es difícil porque el sangrado durante una cirugía está relacionado a factores del paciente como al tipo de cirugía; no se puede imputar por la moda, ni por vecinos cercanos. Adicionalmente la información pronóstica de esta variable está contenida en las variables transfusión y el uso de vasopresores, ya que los pacientes que se transfunden son los que sangran más de 500 cc durante la cirugía y están con soporte vasopresor, por esta razón se elimina la variable sangrada.

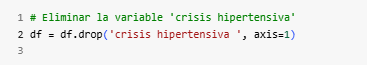


Las variables talla corporal, peso corporal y Creatinina POP¨reemplazar los valores nulos por la mediana.

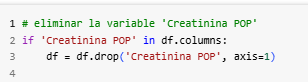


La variable proteinuria preoperatoria con 322 registros indeterminados, se decide eliminar ya que tampoco es imputable fácilmente

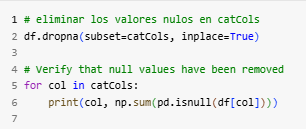
La variable crisis hipertensiva solo tiene 3 registros positivos por lo que se elimina



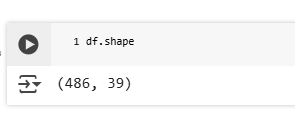
La variable creatinina pop se utiliza para realizar el diagnóstico de lesión renal aguda pop, es decir su significado está ya dentro de esta variable, no aporta al modelo por lo que se elimina



Las variables categóricas se eliminaron los valores nulos, ya que eran pocos registros



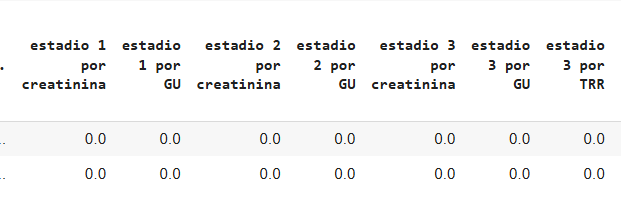
El conjunto de datos final quedó con 39 columnas y 486 registros



**Transformaciones aplicadas**

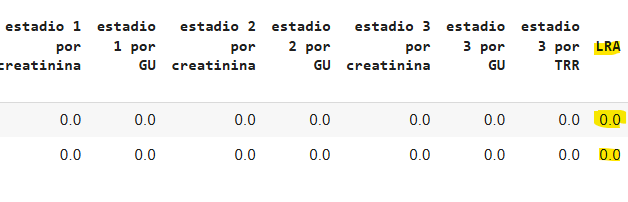
En las variables estadios, el valor Unchecked significa NO ocurrencia del evento y

Checked significa si, realizamos de una vez una dumificacion manual, cambiando Unchecked por 0 y Checked por 1.

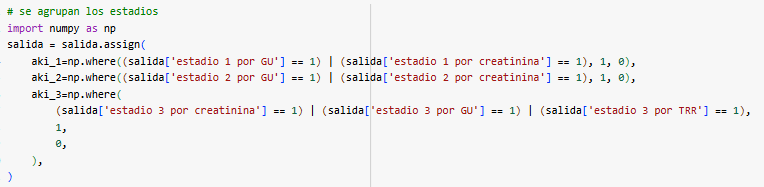


En la variable LRA POP (choice=No presentó), Checked significa que no presento AKI y

Unchecked significa que, si presento AKI, se cambian los valores para mejorar la comprensión Nombre de la variable: LRA Checked por 0 y Unchecked Por 1



El diagnóstico de AKI se realiza por criterios de aumento de la creatinina o por disminución del volumen urinario (gasto urinario). La severidad de la lesión renal aguda está evidenciada en los estadios 1,2 y 3. Se decide agrupar las variables estadios por creatinina y estadio por gasto urinario en una sola variable, ya que indican el mismo nivel de severidad



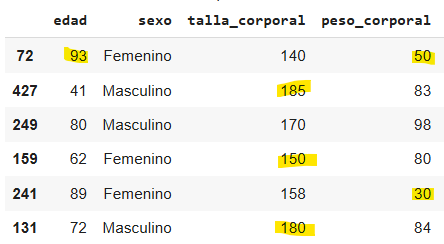


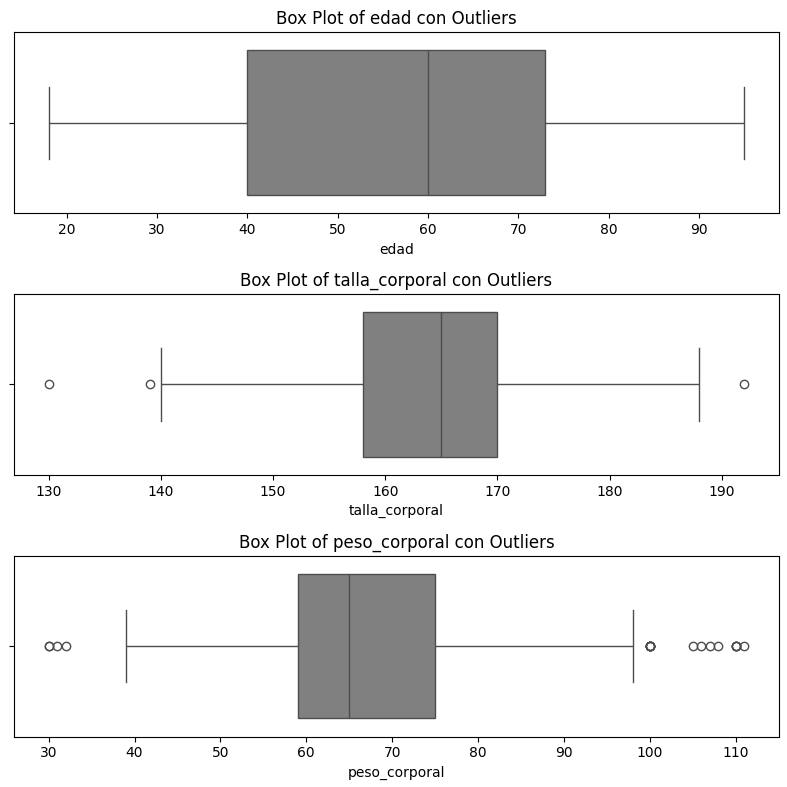
**Valores atípicos**

Se realizó por el método de vecinos cercanos. Al revisar los valores correspondientes en las variables evaluadas; edad, peso y talla, se encontró que eran valores posibles

Ejemplo, un paciente puede tener 93 años de edad, puede parecer atípico, pero no es un dato incorrecto.

Similarmente un paciente con pérdida de peso severa, puede pesar 30 kilos, con referencia a la población si es un valor atípico, pero si es un valor posible para un paciente

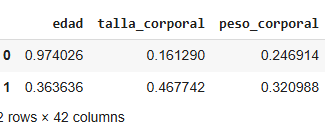


* 

**Normalización**

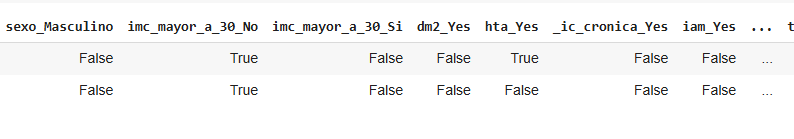
A las variables numéricas se les aplicó Min MaX scaler



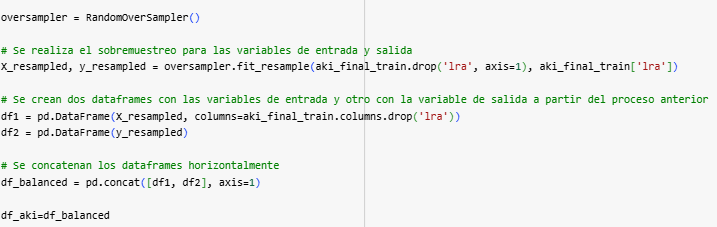


A las variables categóricas se les realizó Get Dummies

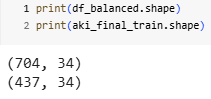




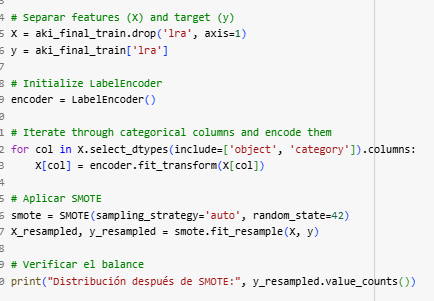
Para tratar de balancear los datos se utilizaron dos técnicas de sobremuestreo:



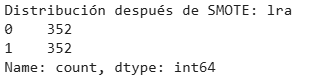
Esta técnica, aunque aumentó el número de datos no corrigió el desbalance:



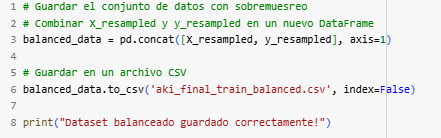
Se utilizó la técnica SMOTE, que se centra en la clase minoritaria



El resultado fue mejor



Este conjunto de datos con mejor balance, se guardó para posteriores análisis



Modelado

Teniendo como base la revisión de la literatura los modelos propuestos son:

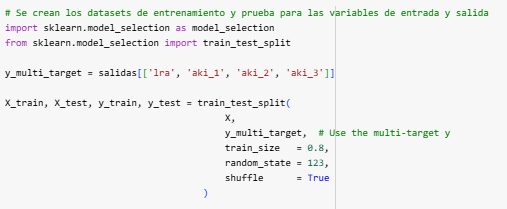
Lineales: regresión Logística

No lineales: Random Forest, árbol de decisiones de aumento de gradiente (GBDT), aumento de gradiente extremo (XGBoost) y Red Neuronal Recurrente

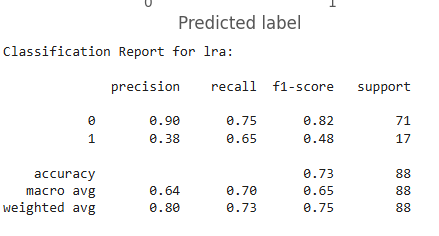
| **Modelo** | **Descripción** | **Ventajas** | **Desventajas** |
| --- | --- | --- | --- |
| Regresión Logística | Estima la probabilidad de un evento basándose en una o más entradas y se utiliza más comúnmente en problemas de clasificación | Cálculo de clasificación más rápido y con menor consumo de tiempo.  Puntuaciones de probabilidad de la muestra de observación intuitivas.  No se ve afectado por la multicolinealidad y se puede combinar con la regularización L2 para resolver el problema.  Bajo coste computacional, fácil de entender e implementar. | El rendimiento computacional se degrada cuando el espacio de características es amplio.  Fácil de subajuste, generalmente poco preciso.  No maneja bien un gran número de características de clase.  Se requiere conversión para características no lineales. |
| Random forest | método de aprendizaje integrado que contiene múltiples árboles de decisión para la clasificación | Los datos de alta dimensión (con muchas características) se pueden calcular sin reducción de dimensionalidad ni selección de características.  Se puede evaluar la importancia de las características.  Se puede evaluar la interacción entre diferentes características.  No se sobreajustan fácilmente.  El entrenamiento es más rápido y es fácil usar métodos paralelos.  Se puede equilibrar el error de conjuntos de datos desequilibrados.  Se puede mantener la precisión si falta una gran parte de las características. | Sobreajuste en algunos problemas de clasificación o regresión ruidosos.  Para datos con atributos con diferentes valores, los atributos con más divisiones de valores tendrán un mayor impacto en el bosque aleatorio. |
| Gradient  boosting  decision tree | Los árboles potenciados utilizan modelos aditivos y algoritmos de avance gradual para implementar el proceso de optimización del aprendizaje | Se puede obtener una alta precisión con un tiempo de ajuste relativamente corto.  Manejo flexible de diversos tipos de datos, incluyendo valores continuos y discretos, para una amplia gama de usos.  Se pueden utilizar algunas funciones de pérdida robustas, que son más robustas a los valores atípicos. | Las dependencias entre aprendices débiles dificultan el entrenamiento de datos en paralelo |
| Extreme  gradient  boosting | El aumento de gradiente extremo (XGBoost) también es un tipo de algoritmo de integración, como un modelo de árbol reforzado, que consiste en una combinación de varios modelos de árbol para formar un clasificador muy robusto. Además, el modelo de árbol utilizado es el modelo de árbol de regresión CART | Es rápido y eficaz en el procesamiento de conjuntos de datos a gran escala y no requiere grandes cantidades de recursos de hardware, como memoria  En comparación con los modelos de aprendizaje profundo, el efecto es similar sin necesidad de ajustar los parámetros.  XGBoost implementa internamente un modelo de árbol potenciado, que puede gestionar automáticamente los valores faltantes | En comparación con el modelo de aprendizaje profundo, no puede modelar la ubicación espaciotemporal ni capturar datos de alta dimensión, como imágenes, voz y texto.  El aprendizaje profundo es mucho más preciso que XGBoost cuando cuenta con una gran cantidad de datos de entrenamiento y puede encontrar un modelo de aprendizaje profundo adecuado |
| Artificial neural  networks | Las redes neuronales artificiales son, en general, similares a redes compuestas por neuronas, que son unidades individuales conectadas entre sí, y cada unidad tiene una cantidad numérica de entradas y salidas, que pueden presentarse en forma de números reales o funciones combinatorias lineales. | Alta precisión de clasificación  Alta capacidad de procesamiento de distribución paralela  Almacenamiento distribuido y alta capacidad de aprendizaje  Gran robustez y tolerancia a fallos ante nervios ruidosos  Capacidad para aproximar completamente relaciones no lineales complejas con función de memoria asociativa | Las redes neuronales requieren una gran cantidad de parámetros, como la topología de la red, los valores iniciales de los pesos y los umbrales.  La imposibilidad de observar el proceso de aprendizaje interno y la dificultad para interpretar los resultados pueden afectar la credibilidad y la aceptabilidad de los resultados.  El tiempo de estudio es demasiado largo y puede que ni siquiera se logre el objetivo del aprendizaje. |

**Regresión logística**

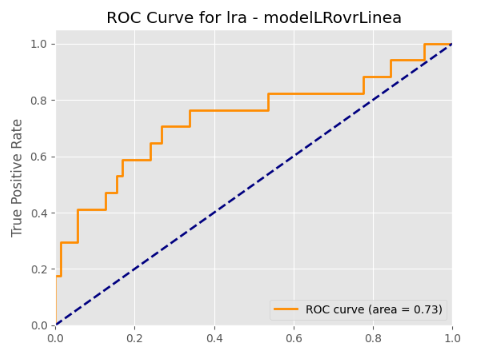
Se realizo una regresión logística para predecir las variables LRA, aki 1, aki 2 y aki 3. El propósito era predecir la ocurrencia de lesión renal aguda y determinar su severidad



El desempeño del modelo se evaluó con una matriz de confusión



También se realizó una curva ROC



El desempeño del modelo no fue bueno para predecir la clase de interes

Las métricas fueron:

Accuracy (Exactitud): 0.74 (74% de predicciones correctas en total).

Clase 0 (Mayoritaria):

Precisión (0.90): Cuando el modelo predice "0", acierta el 90% de las veces.

Recall (0.76): Identifica correctamente el 76% de los casos reales de "0".

F1-Score (0.82): Combinación armónica de precisión y recall (buen equilibrio).

Clase 1 (Minoritaria):

Precisión (0.39): Solo el 39% de las predicciones de "1" son correctas (muchos falsos positivos).

Recall (0.65): Captura el 65% de los casos reales de "1" (pero a costa de baja precisión).

F1-Score (0.49): Bajo debido al trade-off entre precisión y recall.

Problema

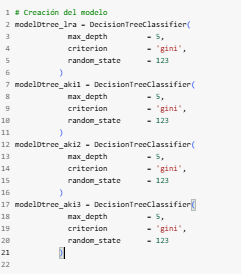
Alto desbalance de clases: La clase "1" tiene solo 17 instancias vs. 71 de la clase "0".

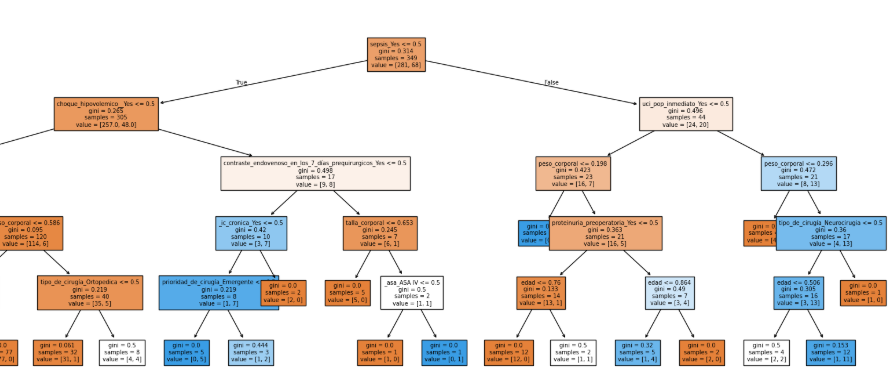
Overfitting hacia la clase mayoritaria: El modelo prioriza aprender "0" porque es más frecuente.

Baja precisión para la clase minoritaria: Predice muchos falsos positivos (ej: clasifica como "1" cuando en realidad es "0").

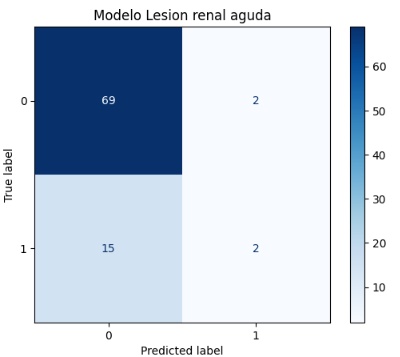
**Arboles de decision**

Se crearon modelos para cada una de las variables objetivos





Se realizo matriz de confusión para determinar el desempeño de los modelos



Para la clase 0 (no LRA), el modelo tiene una precisión alta (0.82), mientras que para la clase 1 (LRA), la precisión es más baja (0.50)

Recall: mide la proporción de verdaderos positivos sobre el total de positivos reales. El modelo tiene un recall muy alto para la clase 0 (0.97), pero bastante bajo para la clase 1 (0.12).

F1-Score: es la media armónica de precisión y recall. Para la clase 0, el F1-Score es alto (0.89), pero para la clase 1 es bajo (0.19)

El modelo predice la clase 0 (no LRA) con alta precisión y recall, pero tiene dificultades para predecir la clase 1 (LRA) con precisión y recall bajo

Se realizaron optimación de hiperparámetros:



Pero no hubo mejoría en el desempeño:

Train accuracy = 100% vs Test accuracy = 32.6%: El modelo memorizó los datos de entrenamiento y no generaliza.

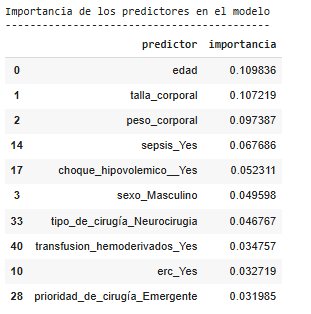
Árbol demasiado complejo (poda insuficiente con ccp\_alpha=0.000001).

Rendimiento pésimo en test:

32.6% de accuracy es peor que un modelo aleatorio (50% en problema binario balanceado).

Alta desviación estándar (11.2%) → Inconsistencia entre folds.

Para finalizar se realizó un análisis de importancia de predictores:



**Random Forest**

Ventajas Random Forest

Reduce overfitting: Al promediar múltiples árboles, suaviza el ruido.

Más estable: Pequeños cambios en los datos no afectan significativamente el modelo.

Maneja mejores datos desbalanceados (con class\_weight='balanced').

No requiere normalización de variables

Se creo un modelo Grid Search basado en out-of-bag score

El Out-of-Bag (OOB) Score es una métrica de evaluación interna que estima el rendimiento de un modelo de Random Forest sin necesidad de un conjunto de prueba separado.

Cada árbol del bosque se entrena con un subconjunto aleatorio de los datos de entrenamiento (muestreo con reemplazo).

Al hacer este muestreo, algunas muestras no son seleccionadas (aproximadamente 37% quedan fuera en cada árbol).

Estas muestras excluidas se llaman "Out-of-Bag" (OOB).



El resultado de los hiperparámetros:

OOB Accuracy: 81.2%

Hiperparámetros óptimos:

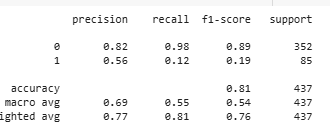
criterion='entropy'

max\_features=9 (de ~39 features, solo 9 se consideran por árbol, limita overfitting).

n\_estimators=250 (número alto de árboles mejora estabilidad).

max\_depth=None (profundidad ilimitada, pero controlada por max\_features).

Las métricas para la predicción de lesión renal aguda fueron:



Clase 0 (No LRA)

Recall = 98%: El modelo identifica casi todos los casos negativos (especifidad alta).

Precisión = 82%: Cuando predice "No LRA", acierta el 82% de las veces.

Clase 1 (LRA**)**

Recall = 12%: Solo detecta 12 de cada 100 pacientes con LRA real.

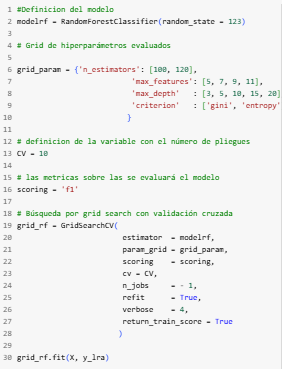
Precisión = 56%: Cuando predice "LRA", solo acierta el 56% de las veces.

Accuracy engañoso (81%)

El alto accuracy se debe a que el modelo prioriza la clase mayoritaria (No LRA).

F1-Score de la clase 1 = 19%: Confirma que el modelo es inútil para predecir LRA.

Se realizo un Grid search con validación cruzada:

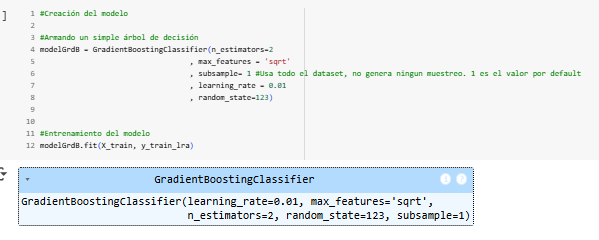


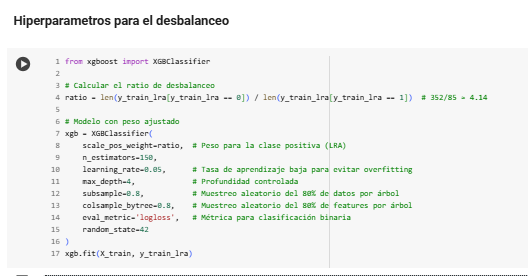
Pero sin mejora en las métricas

**XG Boost**

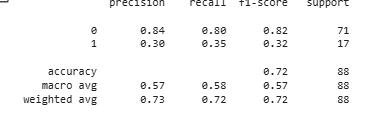
Algoritmo de ensamble por boosting que combina múltiples modelos débiles (generalmente árboles de decisión) de forma secuencial, donde cada nuevo modelo corrige los errores del anterior.

XGBoost optimiza una función de pérdida (ej: para clasificación binaria, usa log-loss o entropía cruzada binaria) que penaliza las predicciones incorrectas.





El resultado:



Precisión 0.30: Solo 30% de los predichos como LRA son verdaderos

Recall 0.35: Solo detecta 35% de los casos reales de LRA (65% no se diagnostican

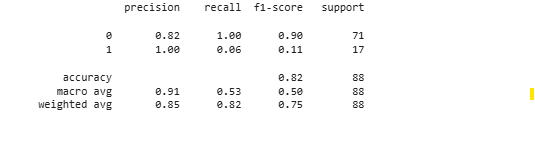
Accuracy (72%): Engañoso, pues refleja el acierto en la clase mayoritaria (No LRA).

El modelo falla en el 65% de los pacientes con LRA

Optimización de los parámetros:



Las métricas:



La exactitud general es del 82%, el recall para la clase 1 (LRA) es extremadamente bajo (6%), lo que significa que el modelo está fallando peligrosamente en detectar pacientes con LRA.

Accuracy engañoso (82%): Se debe a que el modelo está simplemente detectando la clase mayoritaria

Recall: solo detecta el 6% de los casos reales de LRA (94% de falsos negativos)

**Conclusión**

Después de realizar la construcción y evaluación de los diferentes modelos para nosotros es claro que el desbalance de las clases en la variable objetivo está determinando que la predicción sea de la clase mayoritaria.

Debido a este desbalance, cuando se distribuye los datos entre los diferentes estadios de aki, quedan muy pocas instancias para poder hacer alguna predicción, por esta razón los próximos análisis se centrarán solo en la variable LRA (lesión renal aguda).

La solución al pobre desempeño de los modelos no está en la optimización de hiperparámetros como queda demostrado por las diferentes aproximaciones utilizadas; sino en utilizar técnicas que mejoren el balance de los datos.

Ya se cuenta con un conjunto de datos balanceado mediante la técnica SMOTE para el próximo análisis .