

**IDENTIFICACIÓN TEMPRANA Y PREDICCIÓN DE SEPSIS
EN PACIENTES DE UCI A PARTIR DE DATOS
FISIOLÓGICOS**

Modalidad: Exploratorio

PABLO GONZALEZ MESA

**Trabajo de grado para optar al título de
Ingeniería de Sistemas y Computación**

Isis Bonet Cruz, Ph.D

Javier Enrique Camacho Cogollo, MSc

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.



UNIVERSIDAD EIA

INGENIERÍA DE SISTEMAS Y COMPUTACIÓN

ENVIGADO

2020

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

AGRADECIMIENTOS

Quiero darle un agradecimiento muy especial a la profesora Isis Bonet Cruz por despertar mi interés por temas de inteligencia artificial y por las importantes asesorías y contribuciones a este trabajo.

También un agradecimiento muy especial al profesor Javier Camacho por servir de guía para esta tesis y por las importantes contribuciones a este trabajo.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

CONTENIDO

	pág.
INTRODUCCIÓN	11
1. PRELIMINARES	14
1.1 Planteamiento del problema.....	14
1.1.1 Contexto.....	14
1.1.2 Pregunta de Investigación	17
1.2 Objetivos del proyecto	17
1.2.1 Objetivo General.....	17
1.2.2 Objetivos Específicos.....	17
1.3 Marco de referencia	18
1.3.1 Antecedentes	18
1.3.2 Machine learning.....	20
1.3.3 Modelos de <i>machine learning</i>	22
1.3.4 Métodos de validación	28
1.3.5 Métricas.....	29
2. METODOLOGÍA	32

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

2.1	Enfoque general	32
2.2	Datos	33
2.3	Selección de pacientes y recolección de datos en MIMIC III.....	34
2.4	Selección de Atributos	36
2.5	Propuesta de modelo para predicción de sepsis	37
2.5.1	Identificación temprana.....	37
2.5.2	Predicción de sepsis	39
2.6	Selección de pacientes en la institución de salud local	40
2.6.1	Identificación inicio de infección.....	41
2.6.2	Identificación inicio de disfunción orgánica.....	41
2.7	Recolección de datos y Adaptación del modelo a datos de la institución de salud local ...	43
3.	PRESENTACIÓN Y DISCUSIÓN DE RESULTADOS	46
4.	CONCLUSIONES Y CONSIDERACIONES FINALES.....	55
	REFERENCIAS.....	57

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

LISTA DE TABLAS

Tabla 1. Resumen de resultados validación cruzada.....	46
Tabla 2. Resumen de resultados conjunto de validación	48
Tabla 3. Resumen de resultados conjunto de validación más ajuste del umbral	48
Tabla 4. Resumen de resultados en la institución de salud local	54

pág.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

LISTA DE FIGURAS

Figura 1. Cálculo de indicador SOFA (tomada de (Singer et al., 2016)).....	16
Figura 2. Ejemplo algoritmo DTW	23
Figura 3. Matriz distancia	24
Figura 4. Camino más corto.....	25
Figura 5. Representación validación cruzada	29
Figura 6. Matriz de confusión con valores.....	30
Figura 7. Matriz de confusión.....	30
Figura 8. Parámetros ventana de tiempo	37
Figura 9. Diagrama del multclasificador	38
Figura 10. Número de instancias vs parámetro t.....	39

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

LISTA DE ANEXOS

Anexo A. Variables recolectadas de pacientes	61
--	----

pág.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

RESUMEN

Sepsis se define como una disfunción orgánica causada por una respuesta desregulada a una infección por parte del paciente. Es una de las causas de muerte más común en pacientes de unidades de cuidados intensivos en todo el mundo, lo que lo convierte en una problemática de gran importancia. La detección temprana de este síndrome es de gran importancia a la hora de tratar a los pacientes efectivamente.

En este trabajo se propone un modelo de inteligencia artificial para la identificación temprana y predicción de sepsis utilizando la base de datos MIMIC y adaptando los resultados a los datos de una institución local de salud. Se realizaron pruebas con múltiples modelos y parámetros de estos. El mejor resultado para la identificación temprana de sepsis fue el resultado de hacer un modelo multclasificador con diferentes rasgos para los clasificadores, en este caso se usaron un kNN y un XGBoost y utilizar sus resultados como las entradas de una regresión logística. Éste tuvo un AUROC de 0.944. Para predicción de sepsis se propuso un modelo XGBoost capaz de predecir 1, 2 y 3 horas de anticipación con métricas de 0.918, 0.912 y 0.908 respectivamente.

Palabras Claves: Sepsis, Inteligencia Artificial

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

ABSTRACT

Sepsis is defined as an organic disfunction due to an unregulated response to an infection. Sepsis is one of the main causes of death in intensive care units (ICU) all around the world, which makes it a very important problem to tackle. Early sepsis detection plays a big role in taking care of patients effectively.

In this work, an artificial intelligence model for the early identification and prediction of sepsis based in the public database MIMIC is proposed. The model is also adapted to the data of a local health institution. Multiple models with different parameters were tested. A model using an ensemble of classifiers with a XGBoost model and a kNN model and using their results in a logistic regression as meta-classifier obtained an AUROC of 0.944 for early sepsis identification. For sepsis prediction a simple XGBoost model was proposed and was able to predict 1, 2 and 3 hours before the sepsis onset with an AUROC of 0.918, 0.912 and 0.908, respectively.

Keywords: Sepsis, Artificial Intelligence

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

INTRODUCCIÓN

Actualmente las tasas de mortalidad en las unidades de cuidados intensivos (UCI) a causa de sepsis son bastante altas, a pesar de que este síndrome no es la causa principal de entrada de los pacientes a UCI. La detección temprana de este síndrome es una pieza clave a la hora de abordar este problema.

En la primera sección del trabajo se hará una contextualización profunda sobre el síndrome sepsis y porque es actualmente un problema importante para las UCI. Aquí se incluyen cifras que muestran el significativo costo en vidas humanas y en recursos económicos del problema. Más adelante se limita el problema y se define como se va a abordar, planteando un objetivo general y varios objetivos específicos, donde se define que se va a trabajar con métodos de inteligencia artificial para intentar predecir la sepsis con base en datos fisiológicos de los pacientes. Adicionalmente se va a hacer una recopilación de los autores que ha trabajado en este problema, donde se explica que aportes ha dado cada uno y que tan bueno fue el desempeño de los modelos propuestos por ellos.

En la segunda sección del trabajo se explica todo lo que tiene que ver con la metodología del trabajo, que es la sección más robusta y se presenta todo lo que se hizo para abordar el problema. Primero se da una contextualización general a los métodos de *machine learning* que se van a utilizar en el proyecto donde se explica en detalle la estructura de un proyecto de *machine learning*. Luego se explican los procedimientos necesarios para abordar este proyecto específico.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Esto último incluye presentar las fuentes de datos que se van a utilizar para abordar el problema, una introducción a la base de datos MIMIC y a la de la institución de salud local, definir los métodos de validación para los modelos que se van a proponer, describir las métricas que se van a utilizar para medir el desempeño y la descripción de los modelos y otros algoritmos que se van a utilizar para resolver el problema.

Posteriormente se presentan los criterios para escoger el conjunto de pacientes que se va a utilizar para entrenar los modelos, donde se explican los pacientes que se deben excluir y se hace una breve descripción del conjunto obtenido. Para cada uno de los pacientes se identifica si este desarrollo sepsis durante su estancia en UCI y el tiempo en el que este la desarrollo en caso de que aplique. También se explica el procedimiento utilizado para encontrarlo.

Después de esto se va a definir los atributos que se van a utilizar para el modelo y como deben ser recolectados para cada paciente, en estas se tiene una gran cantidad de variables estáticas y de datos en series de tiempo. Lo más importante de esta sección, es que se debe definir claramente la ventana de tiempo en la que se van a recolectar estos datos. Esto determinará si el modelo es de identificación temprana de sepsis o de predicción, y en este último caso también determinará el tiempo anticipación de la predicción.

Más adelante se presentarán los modelos propuestos para dar solución al problema de predicción de la sepsis en la base de datos MIMIC. Aquí se incluye la estructura de los modelos propuestos, las pruebas realizadas variando parámetros y los tipos de modelos y los datos que se utilizaron para realizar el entrenamiento y validación. Este proceso se hace tanto para el problema de identificación temprana como para el de predicción de sepsis.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Luego se explica cómo aplicar todo este proceso para los datos de la institución local de salud. Incluyendo el proceso para obtener a través de consultas en su base de datos el conjunto de pacientes, detectar si estos desarrollaron sepsis y el tiempo en el que la desarrollaron, obtener los atributos que el modelo va a utilizar para cada paciente y la adaptación del modelo para predicción de sepsis para estos datos.

Por último, se presentan los resultados obtenidos para cada uno de los modelos y con base en esto se obtienen conclusiones del trabajo. Aquí se evalúan que tan bueno fue el desempeño de los modelos propuestos y se resaltan los resultados destacados que se obtuvieron. También se hace una reflexión sobre las implicaciones de este trabajo y algunas sugerencias sobre lo que en que se puede trabajar en esta problemática en el futuro.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

1. PRELIMINARES

1.1 PLANTEAMIENTO DEL PROBLEMA

1.1.1 Contexto

Uno de los grandes problemas que tienen actualmente las Unidades de Cuidados Intensivos (UCI) de los hospitales, son las significantes tasas de mortalidad que tienen los pacientes a causa de Sepsis o Shock Séptico, a pesar de que éstas no son las causas principales de entrada a la UCI. Las tasas de ocurrencia de Sepsis en pacientes de UCI son considerables en todas las regiones del mundo, éstas varían desde 13.6% hasta 39.3% según la región (Yasser Sakr et al., 2018). Lo cual implica que incluso en las regiones de menos ocurrencia, siguen teniendo una presencia significativa.

También se ha podido estimar una tasa de mortalidad a nivel mundial de Sepsis en UCI y en hospitales de 25.8% y 35.3% respectivamente. Sin embargo, en diferentes regiones puede llegar hasta 39.5% y 47.2% respectivamente. Para el caso de América del sur se han estimado tasa de mortalidad de 36.3% y 45.4% respectivamente (Yasser Sakr et al., 2018).

También se estima que en 2011 el costo del tratamiento de la sepsis en Estados Unidos sumó 20 mil millones de dólares, que fue aproximadamente un 5.2% del gasto de los hospitales en ese año. Además, se espera que el número de incidencias de sepsis aumente en los próximos años debido a factores cambiantes de la población como su edad, y aumento en comorbilidades (Singer et al., 2016). Todo esto demuestra la severidad de la problemática.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

A pesar de que existen relativamente pocos estudios realizados en Colombia de pacientes de sepsis, se han podido recolectar estadísticas generales sobre el lugar donde los pacientes desarrollan sepsis. A partir de un estudio de 826 personas que desarrollaron sepsis y que fueron ingresadas a cuidados intensivos, se estima que el 51% la desarrollaron en la comunidad, 5% durante la hospitalización general y 44% en UCI. También se observa que el 30.6% de la cohorte presenta afectación de algún órgano. Lo cual evidencia que una gran cantidad de los pacientes (por lo menos a nivel local), desarrollan sepsis en la UCI (Guillermo Ortíz et al., 2014).

La detección temprana de sepsis es de suma importancia para la probabilidad de supervivencia del paciente, pues permite que se asignen oportunamente recursos médicos para controlar la sepsis. De hecho, hay una relación importante entre la tasa de mortalidad del paciente y el tiempo que se demora en detectar y diagnosticar a éste con sepsis. Al montar sistemas y protocolos para la detección temprana de sepsis se ha evidenciado reducciones significantes de la tasa de mortalidad (Westphal et al., 2011).

Según el tercer consenso internacional de definiciones para sepsis y shock séptico, se debe definir a la sepsis como una disfunción orgánica causada por una respuesta desregulada a una infección por parte del paciente (Singer et al., 2016). Operacionalmente una disfunción orgánica se puede definir como el aumento de más de dos puntos en el indicador SOFA.

El indicador SOFA, *Sequential Organ Failure Assessment* por sus iniciales en inglés, es una puntuación que se le asigna a cada paciente según algunas mediciones de datos fisiológicos ese el indicador más usado para detectar disfunción y falla orgánica. El aumento de dos puntos o más este está asociado con una mortalidad dentro del hospital del 10% (Singer et al., 2016).

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Figura 1. Cálculo de indicador SOFA (tomada de (Singer et al., 2016))

Table 1. Sequential [Sepsis-Related] Organ Failure Assessment Score^a

System	Score				
	0	1	2	3	4
Respiration					
Pao ₂ /Fio ₂ , mm Hg (kPa)	≥400 (53.3)	<400 (53.3)	<300 (40)	<200 (26.7) with respiratory support	<100 (13.3) with respiratory support
Coagulation					
Platelets, ×10 ³ /μL	≥150	<150	<100	<50	<20
Liver					
Bilirubin, mg/dL (μmol/L)	<1.2 (20)	1.2-1.9 (20-32)	2.0-5.9 (33-101)	6.0-11.9 (102-204)	>12.0 (204)
Cardiovascular					
	MAP ≥70 mm Hg	MAP <70 mm Hg	Dopamine <5 or dobutamine (any dose) ^b	Dopamine 5.1-15 or epinephrine ≤0.1 or norepinephrine ≤0.1 ^b	Dopamine >15 or epinephrine >0.1 or norepinephrine >0.1 ^b
Central nervous system					
Glasgow Coma Scale score ^c	15	13-14	10-12	6-9	<6
Renal					
Creatinine, mg/dL (μmol/L)	<1.2 (110)	1.2-1.9 (110-170)	2.0-3.4 (171-299)	3.5-4.9 (300-440)	>5.0 (440)
Urine output, mL/d				<500	<200

En la Figura 1 se puede observar como se mide operacionalmente el indicador SOFA, donde el resultado es la suma de los puntajes otorgados por cada una de las variables.

Shock Séptico es definido como un subconjunto de los casos de sepsis en los cuales hay anomalías circulatorias, celulares y metabólicas que se asocian con un aún mayor riesgo de mortalidad. Estos pacientes pueden ser clínicamente identificados por la necesidad de uso de vasopresores para mantener la presión arterial promedio en 65 mm Hg (Singer et al., 2016).

La unidad de cuidados intensivos es un sistema organizado para proveer cuidados a pacientes en estado crítico, que incluyen cuidados médicos intensivos y especializados al igual que la supervisión por parte de enfermeras. (Marshall JC et al. 2017)

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

1.1.2 Pregunta de Investigación

¿Se puede predecir o identificar tempranamente la sepsis en pacientes de UCI utilizando datos fisiológicos de éstos y métodos de inteligencia artificial?

1.2 OBJETIVOS DEL PROYECTO

1.2.1 Objetivo General

Proponer un modelo que permita la predicción temprana de sepsis en pacientes de UCI, utilizando técnicas de inteligencia artificial y datos fisiológicos históricos de bases de datos internacionales y datos de una institución de salud local.

1.2.2 Objetivos Específicos

- Seleccionar los atributos relevantes para describir los pacientes.
- Seleccionar e implementar una función de distancia para las señales multivariantes de mediciones de los pacientes.
- Entrenamiento y pruebas de diferentes modelos con la base de datos MIMIC.
- Selección del mejor modelo para la predicción de sepsis.
- Adaptación del modelo para datos de la institución de salud local.
- Validación del modelo con los datos de la institución de salud local.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

1.3 MARCO DE REFERENCIA

1.3.1 Antecedentes

El estudio de sepsis comienza cuando se le da su primera definición formal al síndrome en el año 1991. A partir de este punto se ha redefinido 2 veces hasta la definición que se utiliza actualmente y que fue adoptada en el 2016. En esta se define sepsis como una disfunción orgánica causada por una respuesta desregulada a una infección por parte del paciente (Singer et al., 2016). También se definen claramente los criterios clínicos para sepsis como una sospecha o documentación de infección y un incremento de dos puntos en el indicador SOFA (Gyawali, Ramakrishna, & Dhamoon, 2019).

Para abordar el problema de identificación de sepsis se han propuesto modelos utilizando *machine learning* para hacerlo. En 2012, ya se habían propuesto modelos basados en redes bayesianas dinámicas, que mostraron una mejora significativa en la identificación de sepsis. Se midió el funcionamiento el modelo con información del paciente durante 3, 6, 12 y 24 horas después de su admisión y se obtuvo un AUROC de 0.911, 0.915, 0.937 y 0.944 respectivamente que es bastante mejor en comparación que solo utilizar los indicadores (Nachimuthu & Haug, 2012). Sin embargo, este es un modelo exclusivamente para la identificación de sepsis y no para su predicción.

En 2017 se propuso un modelo basado en series de tiempo de presión arterial y de frecuencia cardíaca para una detección temprana de sepsis. Aquí fue necesaria la utilización de una buena resolución temporal de datos. Como resultado se obtuvieron dos indicadores que sirven para predecir su ocurrencia bajo el consenso de definición sepsis 3, uno basado en frecuencia cardíaca y otro en presión arterial. También se comprobó que son independientes entre sí (Shashikumar et al., 2017). La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

2017). A partir de estos y datos complementarios se propuso un modelo de detección temprana de sepsis con un AUROC de 0.78.

Actualmente los sistemas para predecir sepsis utilizan herramientas computacionales complejas y pueden predecir con anticipación la sepsis en pacientes de UCI. En particular se han realizado estudios con técnicas de aprendizaje profundo que han logrado resultados que actualmente se consideran estado del arte en cuanto a predicción de sepsis. Se logró predecir la sepsis con una anticipación siete (7) horas antes de su comienzo. Para este resultado se obtuvo una métrica AUPRC (*area under precision recall curve*) de 0.35/0.4 (Moor, Horn, Rieck, Roqueiro, & Borgwardt, 2019). Los modelos basados en aprendizaje profundo tienen sin embargo una desventaja y es que no pueden ser explicados en términos humanos fácilmente, cosa que para esta problemática puede ser importante.

A pesar de todos los avances, todavía es un problema abierto para su estudio y existen muchos caminos para abordarlo que vale la pena explorar. En particular todavía es importante explorar lo que es la explicabilidad de los modelos, si es posible predecir el síndrome con datos más sencillos, su aplicación a diferentes demografías, su aplicación a datos de UCI en tiempo real, etc.

En los últimos años se ha enfocado en abordar el problema de la sepsis a través de las herramientas de *machine learning* obteniendo avances importantes para su solución.

Por ejemplo, en 2018 se propuso un algoritmo predictivo llamado *InSight* que utiliza únicamente datos de signos vitales del paciente. Este obtuvo resultados muy satisfactorios, permitiendo predecir las ocurrencias de sepsis, sepsis severa y shock séptico con 4 horas de anticipación y con un AUROC de 0.92, 0.85 y 0.96 respectivamente (Mao et al., 2018). A pesar de

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

estos resultados prometedores, es importante tener en cuenta que el modelo implementado obtiene se basa en la información codificada en ICD-9 de los hospitales para determinar que es sepsis y que no lo es, lo cual es un código diseñado para operaciones administrativas en los hospitales y no necesariamente para fines médicos (Mao et al., 2018).

Más adelante se publica una investigación que estudia un modelo también basado en signos vitales para predecir la sepsis con mucho más tiempo de anticipación. Este se evaluó prediciendo la sepsis 24 y 48 horas antes de su ocurrencia. Para estos se obtuvo un AUROC de 0.84 y 0.83 respectivamente, ampliando considerablemente el tiempo con el que se puede predecir la sepsis (Barton et al., 2019). A pesar de esto el algoritmo no ha sido probado en un hospital en tiempo real, lo cual puede afectar su desempeño significativamente (Barton et al., 2019).

Actualmente, el profesor Javier Camacho de la Universidad EIA se encuentra realizando su tesis doctoral en el tema de predicción anticipada de sepsis en las unidades de cuidados intensivos. Donde se busca implementar sistemas de detección temprana de la sepsis específicamente para el caso de Colombia, actualmente en la institución de salud local. De ahí surge esta tesis como apoyo a su tesis doctoral. Se busca entonces un modelo basado en información de fácil acceso capaz de alimentarse de los datos de los pacientes y predecir la sepsis en tiempo real con un buen desempeño.

1.3.2 Machine learning

Recientemente con el boom de las tecnologías de la información, se ha abierto la posibilidad de utilizar grandes cantidades de datos de pacientes para modelar y predecir el comportamiento de su salud. En particular para abordar la problemática de la Sepsis se han utilizado herramientas de análisis estadístico, de *machine learning* y de otras herramientas computacionales, La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

logrando grandes avances en este. Para más claridad, es necesaria una definición precisa de *Machine learning* y de algunos otros conceptos de las ciencias de la computación que se utilizarán durante el trabajo, al igual que de herramientas específicas que se utilizarán para abordar el problema.

Machine learning (o aprendizaje automático) es la rama de las ciencias de computación que estudia métodos con los cuales un computador puede aprender a través de la experiencia o datos recopilados (Michie, Donald, 1968). Cuando un problema es muy complejo para resolver, el uso de métodos de *machine learning* es de gran ayuda, permitiéndole a la maquina resolver el problema a partir de lo aprendido sin el limitante de la capacidad del programador. Gracias a los avances en la velocidad de los procesadores y en la cantidad de datos disponibles, esta rama ha cogido bastante fuerza en los últimos años.

Dentro de esta rama existen métodos de aprendizaje supervisado que utilizan datos con su solución deseada para aprender el comportamiento de estos y generalizar este resultado para nuevos datos. En estos problemas los datos consisten en atributos (*features*) y de una etiqueta o atributo objetivo (*target* o *label*). Los atributos son las variables de entrada y la etiqueta es la variable de salida del modelo de machine learning (Géron, 2019).

Dentro de los problemas de aprendizaje supervisado se encuentran los problemas de clasificación que tienen como objetivo clasificar una instancia de los datos en una de varias clases, es decir que la etiqueta sea una variable discreta que solo puede tomar una cantidad finita de valores. Un ejemplo de esto puede ser determinar un correo electrónico como *spam* o no a partir de información del correo (Géron, 2019). En este caso cada instancia se clasifica en una de dos clases

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

por lo que este problema se considera de clasificación binaria. Para estos usualmente también se utiliza que el modelo tenga como salida probabilidades indicando la pertenencia de una instancia de los datos a cada una de las clases. Precisamente, para el caso de este trabajo, el problema de predicción de sepsis se puede plantear como uno de clasificación binaria, donde se quiere predecir para cada paciente si tendrá sepsis o no.

Una característica de algunos de los datos de problemas de *machine learning*, es que pueden ser balanceados o desbalanceados. Los conjuntos balanceados tienen la propiedad de que la variable objetivo se distribuye aproximadamente de forma uniforme sobre todos los valores que puede tomar. Por ejemplo, para el caso de un problema de clasificación binaria un conjunto balanceado se tiene aproximadamente el mismo número de instancias de la primera clase como de la segunda. En los conjuntos desbalanceados, la variable objetivo no se distribuye uniformemente. Se abordará este problema como un problema de *machine learning* donde al final se presentará un modelo capaz de predecir la sepsis en pacientes en UCI. A continuación, se describen los datos y los métodos de *machine learning* utilizados.

1.3.3 Modelos de *machine learning*

K vecinos más cercanos

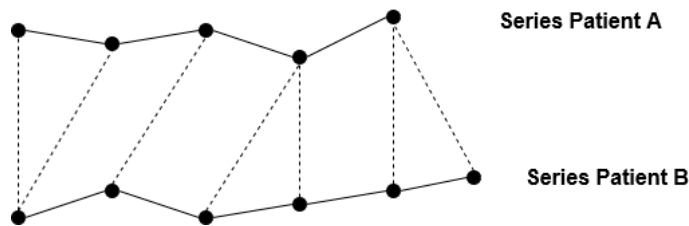
El modelo de K vecinos más cercanos (kNN, de las siglas en inglés k-Nearest Neighbors) es un modelo de clasificación de basado en casos, es decir utiliza todos los datos de entrenamiento para clasificar. Funciona definiendo una medida de similitud entre diferentes instancias, usualmente se utiliza la distancia euclidiana. Entrenar el modelo es bastante simple, únicamente se debe almacenar en memoria el conjunto de datos de entrenamiento. Para clasificar una nueva instancia

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

nunca vista se utiliza un parámetro entero k y se calcula la similitud de ésta con todos los datos del conjunto de entrenamiento. Luego se obtienen las k instancias más cercanas a ésta. La clase de esta nueva instancia será la que más se repita en los vecinos cercanos. También se pueden calcular probabilidades de que la instancia pertenezca a cada una de las clases como el número de instancias similares a esta de la clase sobre k .

En este trabajo se utilizó un modelo kNN con los datos de series de tiempo. Para esto se requiere un método para comparar series de tiempo, aunque sean de distinto tamaño. Se decidió utilizar el algoritmo *Dinamic Time Warping* (DTW), que es un método más flexible para hacer comparaciones entre series y ha dado buenos resultados al ser utilizado en la clasificación de pacientes de sepsis (Moor et al., 2019).

Figura 2. Ejemplo algoritmo DTW



Para entender el funcionamiento de este algoritmo se va a mostrar un ejemplo de como comparar dos series de tiempo. Suponga que se quieren comparar las series $a=[8 \ 12 \ 13 \ 14 \ 10]$ y $b=[10 \ 12 \ 14 \ 13 \ 12 \ 9 \ 7]$. En primer lugar, se debe construir una matriz de diferencias de la siguiente manera.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Figura 3. Matriz distancia

2	4	6	5	4	1	1
2	0	2	1	0	3	5
3	1	1	0	1	4	6
4	2	0	1	2	5	7
0	2	4	3	2	1	3

$$d_{i,j} = |a_i - b_j|$$

La matriz resultante tendrá un número de filas igual a la longitud de a y un número de columnas igual a la longitud de b . Cada una de las entradas de la matriz es la diferencia entre el elemento i de la secuencia a y el elemento j de la secuencia b , según su fila y columna respectivamente. El algoritmo DTW define la distancia entre a y b como el camino con el menor valor entre la esquina superior izquierda y la esquina inferior derecha utilizando únicamente movimientos horizontales hacia la derecha, verticales hacia abajo y diagonal hacia abajo y a la derecha. El valor del camino es la suma de todos los elementos por los que pasa.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Figura 4. Camino más corto

2	4	6	5	4	1	1
2	0	2	1	0	3	5
3	1	1	0	1	4	6
4	2	0	1	2	5	7
0	2	4	3	2	1	3

En la ilustración 6 se ve el menor camino para la matriz construida. Para adaptar este algoritmo a la comparación de signos vitales, este debe poder manejar series multivariadas. Para esto lo único que se debe cambiar es la diferencia que se utiliza, donde ya no es una diferencia numérica en valor absoluto, sino que es una diferencia vectorial. En este caso se va a utilizar la distancia manhattan con la que se obtiene un buen desempeño.

Regresión Logística

El modelo de Regresión Logística es uno de los modelos más clásicos de clasificación binaria. Este calcula una combinación lineal de los atributos de la instancia y luego este resultado lo ajusta entre 0 y 1. Formalmente utiliza la función $\sigma(x) = \frac{1}{1 + e^{-x}}$ sobre el resultado de la combinación lineal. La salida del modelo se puede interpretar como la probabilidad de que la instancia sea positiva (Peng, Lee, & Ingersoll, 2002).

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Multclasificadores

Un método bastante usado cuando se tienen varios clasificadores es construir un multclasificador o ensamble de clasificadores. Para combinar clasificadores existen varias tendencias, cambiar la entrada de los datos con un mismo modelo entrenado varias veces y combinar su salida con un voto pesado, como los hacen Bagging y Boosting. Otra técnica que usualmente da buenos resultados se llama *stacking* y consiste en utilizar diferentes modelos de clasificadores con los mismos datos de entrenamiento y los resultados de ellos se utilizan para entrenar un metaclasificador que finalmente combina sus salidas (Ranawana & Palade, 2016).

La otra forma de hacer multclasificadores es cambiar los rasgos de entrada para clasificador que se utilice. En este trabajo precisamente, por los diferentes tipos de datos se va a utilizar este tipo de multclasificador.

XGBoost

XGBoost es un modelo de multclasificador que utiliza arboles de decisión y la técnica de *Gradient Boosting* para obtener un clasificador de muy buen desempeño para problemas de clasificación y regresión. La técnica de *Gradient Boosting* consiste en entrenar una gran cantidad de clasificadores consecutivamente, en este caso arboles de decisión, tal que cada clasificador se entrene para corregir los errores del clasificador anterior. XGboost tiene además la ventaja de que maneja valores nulos sin tener que imputarlos (Agarwal, Wadhwa, & Chandra, 1994).

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Otros Modelos

A continuación, se van a presentar brevemente algunos modelos que se van a utilizar como referencia en el trabajo para comparar el desempeño obtenido por los anteriores.

Los modelos de árboles de decisión es un modelo basado en condiciones. Se utiliza el conjunto de entramiento para generar condiciones con base en los atributos y con estas generar la estructura del árbol de decisión. Para clasificar, inicialmente se empieza en el nodo principal del árbol (raíz), luego según la condición que se genera se decide por cual de los nodos hijos continuar y así sucesivamente hasta las hojas del árbol. Según la hoja en que se termine se obtienen probabilidades de pertenencia a cada clase (Géron, 2019).

Los modelos de tipo Bosque Aleatorio (*random forest*) son modelos basados en árboles, en estos se construyen y se entrenan múltiples árboles de decisión variando aleatoriamente algunos de sus parámetros. Luego se construye un ensamble con estos y este se utiliza para hacer una predicción final (Géron, 2019).

Los modelos *Naive Bayes* utilizan probabilidades bayesianas para predecir una probabilidad de pertenencia a cada una de las clases. En estos se asume la independencia entre los diferentes atributos y se utiliza el teorema de *Bayes* para obtener una probabilidad condicional de pertenencia a las clases dados los atributos que tiene cada instancia (Lee et al., 2001).

Los modelos tipo máquina vector soporte son modelos de clasificación binaria que dividen el espacio de los atributos en dos secciones diferentes, cada una correspondiente a una de las clases. Estas buscan el hiperplano que divida de manera más adecuada el espacio, buscando maximizar la distancia entre el hiperplano y las instancias. También es posible utilizar un *kernel* para hacer una

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

transformación no lineal del espacio de los atributos y dividir de la misma manera el espacio resultante (Géron, 2019).

También se propone otra función de distancia que se va a utilizar junto al kNN para ver su desempeño y compáralo con el algoritmo DTW. Esta se va a ser referida como *función desface* y consiste en hacer varias comparaciones vectoriales simples, variando el desface de estas y luego obteniendo el promedio.

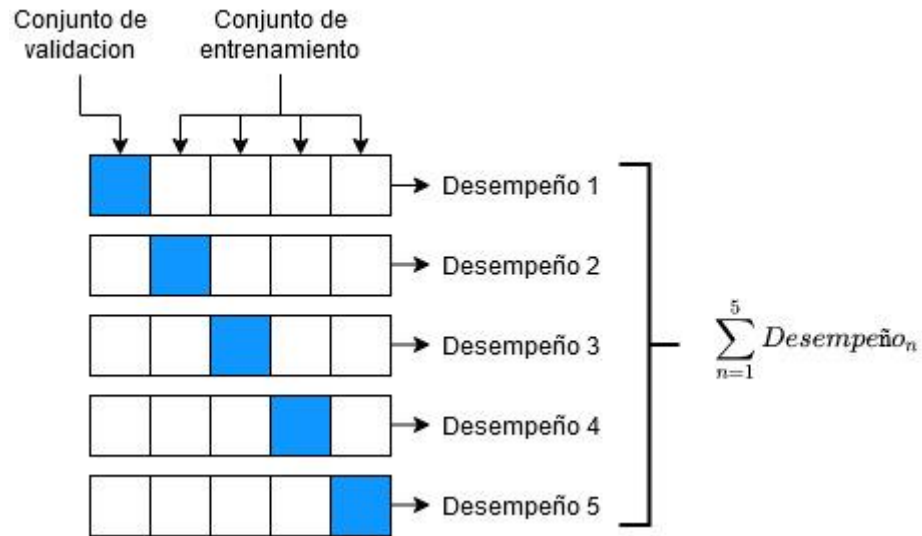
1.3.4 Métodos de validación

Para la implementación de un modelo de *machine learning* es necesario asegurar que el modelo generalice bien a la hora de ser utilizado en datos nunca vistos. Por esto se debe utilizar algún método de validación del modelo. Una primera opción es dividir los datos en un conjunto de datos de entrenamiento y un conjunto de datos de evaluación. El primero es utilizado para el entrenamiento del modelo, luego se utiliza el modelo entrenado para predecir los *targets* del segundo conjunto y se comparan las salidas del modelo con las reales para medir su desempeño.

Otro método más robusto para validar que el modelo generaliza bien para datos fuera del set de entrenamiento es la Validación Cruzada (*k-fold cross-validation*). Consiste en dividir el set de entrenamiento en k partes iguales, entrenar el modelo k veces utilizando $k-1$ de estas partes como conjunto de entrenamiento y la restante como conjunto de evaluación. Luego se promedia el desempeño de todos los entrenamientos.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Figura 5. Representación validación cruzada



La Figura 5 representa una validación cruzada con k=5 pero en este trabajo se utilizará validación cruzada con k=10.

1.3.5 Métricas

Para medir el desempeño del modelo mencionado en la sección anterior es necesario definir las métricas con las que se va a determinar qué tan bueno es el modelo.

Matriz de confusión: Es una matriz donde las columnas representan el valor real la variable, las filas representan como lo clasificó el modelo y cada una de las entradas de la matriz es el número de instancias que corresponden al cruce entre filas y columnas (Géron, 2019). A continuación, se ve un ejemplo de una matriz de confusión para un problema de clasificación binaria:

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Figura 6. Matriz de confusión con valores

	Valor real 1	Valor real 0
Clasificado como 1	50	20
Clasificado como 0	30	100

En la ilustración 2 se observa que cincuenta (50) instancias fueron clasificadas correctamente como pertenecientes a la clase 1, veinte (20) fueron clasificadas en la clase 1 pero pertenecían a la clase 0, treinta (30) fueron clasificadas en la clase 0 pero pertenecían a la clase 1 y cien (100) instancias fueron clasificadas correctamente como pertenecientes a la clase 0.

Figura 7. Matriz de confusión

	Valor real 1	Valor real 0
Clasificado como 1	VP	FP
Clasificado como 0	FN	VN

Para el caso de clasificación binaria, los valores de la matriz de confusión se llaman como se muestra en la ilustración 3, VP (verdaderos positivos), FP (falsos positivos), FN (falsos negativos) y VN (verdaderos positivos) respectivamente. Estas variables se utilizarán para definir las siguientes métricas.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Accuracy: Se calcula como $\frac{VP + VN}{VP + FP + FN + VN}$. Mide la razón entre el número de

instancias clasificadas correctamente por un modelo y el número total de instancias (Géron, 2019).

Tasa de verdaderos positivos (TPR o *Recall*): Se calcula como $\frac{VP}{VP + FN}$. Este valor nos

dice la probabilidad de que una instancia positiva sea identificada correctamente por el modelo (Géron, 2019).

Precision: Se calcula como $\frac{VP}{VP + FP}$. Este valor nos dice la probabilidad de que una

instancia identificada como positiva por el modelo sea en realidad positiva (Géron, 2019)..

Tasa de falsos positivos (FPR): Se calcula como $\frac{FP}{FP + VN}$. Mide la razón entre el número

instancias que fueron identificadas como positivas pero que su valor real es negativo y el total de instancias negativas. (Géron, 2019)

Área bajo la curva COR (AUROC): Se utiliza para los modelos de clasificación binaria que tienen como salida una probabilidad o un puntaje para cada instancia. Es el área bajo la curva resultante de graficar el TPR contra FPR dados muchos umbrales de corte entre clases. Se puede mostrar que esta métrica es la probabilidad de que una instancia “positiva” obtenga una probabilidad o un score mayor a una instancia “negativa” (Cortes & Mohri, 2004).

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

2. METODOLOGÍA

2.1 ENFOQUE GENERAL

Inicialmente hay que hacer una recopilación de los datos a trabajar en el problema. Hay que estudiar varias fuentes de información y cuales datos son relevantes para la variable a predecir, en este caso, si un paciente tiene sepsis o no. En particular para el problema a trabajar se definió que una fuente de información inicial será la base de datos MIMIC pues esta contiene una gran cantidad de pacientes con datos de buena calidad.

Luego es necesario hacer una minería, preprocesamiento y limpieza de los datos para obtener un grupo de estos con los que se va a alimentar el modelo de *machine learning*. Particularmente, es importante obtener en esta etapa una cohorte de los pacientes que van a ser utilizados para entrenar y evaluar el modelo. También se deben definir claramente los criterios con los que se obtuvo esta cohorte de pacientes. Por ejemplo, en varios estudios se eliminan pacientes pediátricos, para tener un grupo independiente de esa variable (Desautels et al., 2016).

Más adelante es importante definir cómo va a ser evaluado el modelo, definir las métricas y los métodos de validación que se van a utilizar para verificar que el modelo tenga un buen desempeño y pueda generalizar bien los resultados.

Luego se deben definir y construir los atributos que va a recibir el modelo como entradas. Es importante que estas tengan una relación estadística, intuitiva o teórica con el problema para que el modelo pueda más tarde capturar la esencia de estos y tener un buen aprendizaje y que pueda

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

generalizar para datos que nunca se hallan visto. Se trabajará con los signos vitales y otros datos fisiológicos de los pacientes como atributos del modelo, puesto que con estos se han obtenido resultados satisfactorios en estudios anteriores y en la base de datos MIMIC se tiene una buena resolución temporal de estos.

Por último, está la etapa de entrenamiento y evaluación del modelo (clasificador) (Desautels et al., 2016) donde el objetivo es utilizar varios modelos con sus algoritmos de aprendizaje respectivos para obtener resultados. Se debe tener también una métrica bien definida ante la cual se va a medir el modelo, en este caso en específico se utilizará AUROC pues es bastante usado en diferentes estudios de predicción de sepsis y es muy buena métrica en general para los problemas de clasificación binaria.

Adicionalmente, después de abarcar el problema en MIMIC, se adaptará lo realizado para los datos de la institución de salud local.

2.2 DATOS

MIMIC es una base de datos abierta al público desarrollada por el laboratorio de fisiología computacional del MIT que comprende datos de la salud asociados a 60.000 admisiones a UCI. Contiene entre otros, signos vitales, exámenes de laboratorio y medicaciones. Se utilizará esta base de datos como fuente para obtener las entradas del modelo. Para este trabajo se necesita obtener un conjunto de admisiones a UCI de pacientes, con la información de si desarrollaron o no sepsis (es posible que un mismo paciente tenga varias admisiones). Para cada admisión se necesita determinar si se desarrolló sepsis dentro de la UCI, el tiempo en el cual se desarrolla (en caso de que aplique) y

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

recolectar las variables fisiológicas del paciente durante su admisión que se utilizarán como atributos del modelo. MIMIC contiene datos de órdenes de cultivos y de órdenes de antibióticos, las cuales se pueden utilizar para determinar si el paciente tuvo alguna sospecha de infección y la hora en que la tuvo. Para detectar una disfunción orgánica en estos pacientes, se puede hacer un cálculo del indicador SOFA hora a hora (las variables necesarias también están contenidas en MIMIC) y detectar cuándo existió un cambio de más de 2 puntos en este indicador. En admisiones que tengan presencia de ambas se marcan como que desarrollaron sepsis y se marca su tiempo de inicio será lo último que ocurrió entre la sospecha de infección y la disfunción.

Por otro lado, también se tiene acceso a una parte de la base de datos de la institución de salud local. Los procesos realizados en MIMIC deben ser replicados en esta tan similar como sea posible. Claramente es inevitable que falten datos de algunas de las variables deseadas, por lo cual estas deberán ser excluidas y trabajar solo con las que sea posible. Afortunadamente, la gran mayoría de las variables se encuentran también en esta base de datos, por lo que no hay que hacer una adaptación muy grande.

2.3 SELECCIÓN DE PACIENTES Y RECOLECCIÓN DE DATOS EN MIMIC III

Junto con el profesor Javier Camacho de la Universidad EIA se hizo la recolección de datos. Basándose en varios artículos del tema se desarrolló la consulta para realizar este trabajo. Para determinar la cohorte de pacientes se excluyeron los pacientes menores de 14 años en el tiempo de su admisión, los que se les detectó que se desarrolló sepsis antes del ingreso al UCI y los que hubieran tenido cirugía cardiovascular. También se incluyeron únicamente los datos clínicos

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

suministrados por el sistema de información clínica MetaVision. Por último, también se excluyó las mediciones que se consideran por fuera de los rangos en los que pueden estar, en las que claramente hubo un error de medición. Todo esto se hizo de acuerdo a lo que varios autores han hecho para excluir pacientes y datos en MIMIC (Cloud, 2018). Una vez aplicados todos estos filtros resultan 11,811 admisiones, de las cuales 2482 tienen sepsis y 9329 no, donde es posible que varias admisiones sean de un mismo paciente.

Para calcular el tiempo de inicio de la sepsis se calculó el tiempo de inicio de infección y el tiempo de inicio de disfunción orgánica para cada registro. Para el tiempo de infección se utiliza el tiempo donde se realizó la primera orden de antibióticos o cultivos para cada admisión. Luego para el tiempo de disfunción se calculó el SOFA hora a hora y se detectó cuando ocurrió un cambio de más de 2 puntos en este.

En el **Anexo A** se puede ver las variables que fueron recolectadas de cada admisión, teniendo en cuenta que es posible que existan valores faltantes. Estas se dividen en datos estáticos como demografías, comorbilidades, datos estadísticos de signos vitales, laboratorios, etc. Y datos de signos vitales medidos como series de tiempo. Específicamente, mediciones durante el tiempo de estancia en UCI de las variables frecuencia cardíaca, presión arterial sistólica, diastólica y promedio, frecuencia respiratoria, glucosa y temperatura.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

2.4 SELECCIÓN DE ATRIBUTOS

Para los datos estáticos se seleccionaron las variables teniendo en cuenta la recomendación de expertos, consulta a distintos artículos publicados sobre este tema y su disponibilidad en las bases de datos que se tienen.

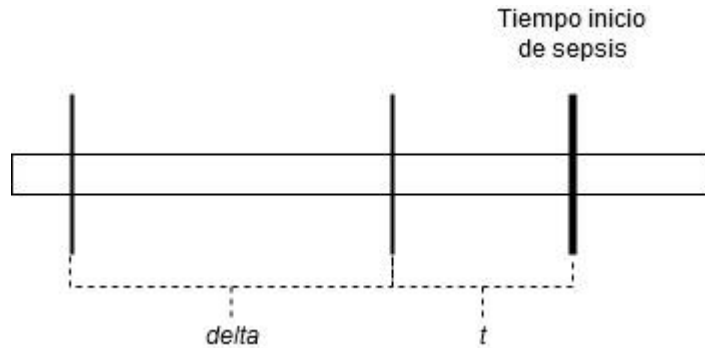
Para los datos de series de tiempo se seleccionaron variables que puedan ser medidas por los equipos de monitoreo de pacientes de UCI. Esto en un futuro puede facilitar la implementación del modelo basado en estos datos para la detección de sepsis y que no se requiera mucho esfuerzo humano adicional. Como anteriormente fue mencionado las variables de signos vitales recolectadas fueron frecuencia cardíaca, presión arterial sistólica, diastólica y promedio, frecuencia respiratoria, glucosa y temperatura. Las mediciones de estas variables fueron agrupadas hora a hora para cada paciente.

A la hora de agregar estos datos al modelo, se debe tener en cuenta el tiempo de inicio de la sepsis. Si agregan datos después de este tiempo, puede mejorar el desempeño de este utilizando los datos propuestos. Sin embargo, a la hora de implementarse en una UCI, es muy probable que su desempeño empeore significativamente debido a que si se intenta predecir la sepsis con algún tiempo de anticipación no se van a tener los datos del paciente después del inicio de la sepsis. Como el modelo fue entrenado por estos datos, si no se tienen estos, probablemente perderá capacidad para predecir. Por lo tanto, se van a limitar los datos de las variables recolectadas con base en dos parámetros. El primer parámetro (t) va a medir el número de horas antes del tiempo de inicio de la sepsis hasta donde se van a tomar datos de signos vitales. El segundo parámetro (δ) es la ventana de tiempo en horas en la cual se van a medir los datos. Por ejemplo, si $t=3$ y $\delta=24$ se

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

terminaría de medir la serie de tiempo y los datos estáticos 3 horas antes del tiempo de inicio de la sepsis y se haría con una ventana 24 horas hacia atrás. Se obtendría entonces 7 series de tiempo de 24 valores cada una (posiblemente con algunos valores faltantes).

Figura 8. Parámetros ventana de tiempo



2.5 PROPUESTA DE MODELO PARA PREDICCIÓN DE SEPSIS

2.5.1 Identificación temprana

Para la identificación temprana de la sepsis se tomaron parámetros t y $delta$ de 0 y 4 para las series de tiempo respectivamente, y para los datos estáticos un t y $delta$ de 0 y 24 respectivamente. Como modelo se propuso un ensamble entre un clasificador X que tenga como entrada los datos estáticos y un kNN que tenga como entrada los datos de las series de tiempo y una función distancia que pueda manejar series de tiempo de distinto tamaño. Las salidas de ambos son utilizadas como las entradas de un metaclassificador como se observa en el siguiente diagrama.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Figura 9. Diagrama del multclasificador



Se probaron diferentes modelos para el clasificador X y para el metaclasificador. Se determinó que los mejores resultados se obtenían con un XGBoost y con una regresión logística respectivamente. El primero permite trabajar con datos nulos sin tener que hacer ninguna imputación. Para la función de distancia del kNN que mejores resultados otorgó fue el algoritmo DTW, por lo que se decidió trabajar con este.

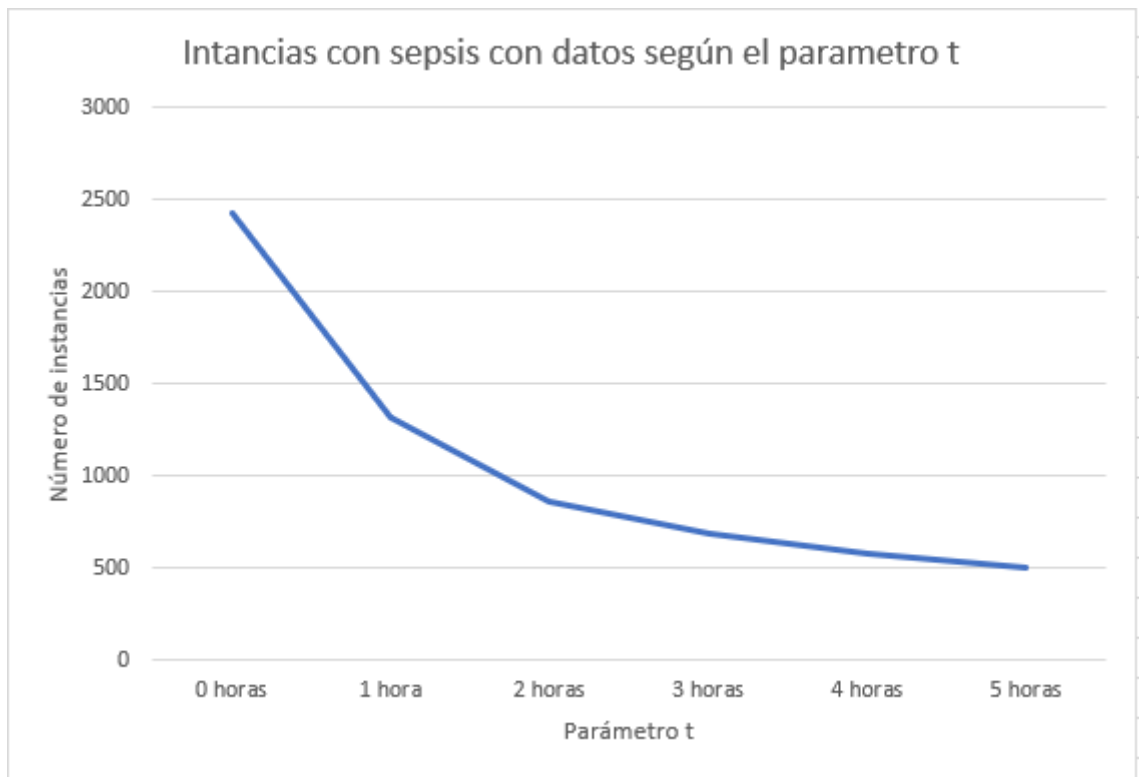
Empíricamente se determinó que entrenar un modelo kNN para este problema con un conjunto de datos desbalanceado afecta en gran medida de forma negativa al desempeño del modelo. Sin embargo, es importante que los resultados que entregue el modelo generalicen bien en un escenario que sea lo más parecido posible a la realidad, donde hay un desbalance de datos. Por lo tanto, de las 11800 admisiones a la UCI se escogió un conjunto de entrenamiento balanceado de 3400 admisiones, 1700 desarrollaron sepsis y 1700 no. Y un conjunto de validación desbalanceado de 3582 admisiones, 782 desarrollaron sepsis y 2800 no. De esta manera se puede entrenar el modelo efectivamente sin afectar su desempeño y asegurar que este puede generalizar correctamente a datos reales de UCI.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

2.5.2 Predicción de sepsis

Para la predicción de sepsis se debe escoger un parámetro t mayor a 0. Esto desafortunadamente es problemático para la recolección de los datos de series de tiempo en MIMIC, debido a que los pacientes tienen una ventana muy corta desde se entra a UCI hasta que tuvo su ocurrencia de sepsis. Entre más grande sea el parámetro t menos datos de series de tiempo y menos instancias se pueden utilizar para entrenar el modelo.

Figura 10. Número de instancias vs parámetro t



En la Figura 10 se observa como el número de instancias que desarrollaron sepsis y que tienen al menos un dato al cambiar el parámetro t . Solo se va a utilizar los datos de series de La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

tiempo con $t=1$, debido a la poca cantidad de instancias que se pueden usar con un t más grande. Por ejemplo, con $t=2$ se tienen únicamente 864 instancias con sepsis y 181 de estas tienen apenas un dato.

En este caso se va a implementar un multclasificador similar al de la Figura 9. Se va a utilizar un conjunto de entrenamiento balanceado con 900 instancias con sepsis y 900 sin sepsis y un conjunto de validación con 421 instancias con sepsis y 1604 instancias que no lo desarrollaron.

También es posible variar más el parámetro t y utilizar únicamente los datos estáticos de los pacientes, donde se utiliza un único clasificador. Se van a utilizar los valores para el parámetro t de 1, 2 y 3, es decir, se va a probar distintos tiempos de anticipación para la predicción. El parámetro δ se va a fijar en $24-t$. Para mantener el estándar de la sección anterior se escogieron exactamente el mismo conjunto de entrenamiento y el mismo conjunto de validación. Es decir, un conjunto balanceado de 3400 admisiones y uno desbalanceado de 3582 admisiones respectivamente. De esta manera las métricas de este modelo pueden ser comparadas directamente con las de identificación temprana de sepsis.

2.6 SELECCIÓN DE PACIENTES EN LA INSTITUCIÓN DE SALUD LOCAL

La institución de salud local nos brindó acceso a una base de datos donde se tiene una gran cantidad de variables fisiológicas, ordenes de cultivos, ordenes de antibióticos, demografías, comorbilidades, entre otros. También se tiene acceso a una tabla de admisiones desde 2016 a 2019, con 4445 registros diferentes, de los cuales 704 se les diagnosticó sepsis. Es posible que diferentes

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

admisiones pertenezcan a un mismo paciente, pero que se diferencien por el tiempo en que ocurrió la admisión.

Sin embargo, no se tiene identificado el tiempo en el cual empezó la sepsis para cada registro. Por lo tanto, se debe realizar un proceso de identificación de los tiempos de inicio de la sepsis. Se necesita entonces calcular un tiempo de inicio de infección y un tiempo de inicio de disfunción orgánica. Para esto se realizará el proceso que se hizo para obtenerlos en la base de datos de MIMIC, pero adaptándolo a la base de datos de la clínica. Todos los procedimientos en esta base de datos se realizaron en consultas SQL.

2.6.1 Identificación inicio de infección

Para la identificación de inicio de infección se puede detectar a través de las ordenes de antibióticos y de cultivos. Se hará entonces una consulta a la base de datos donde a cada una de las admisiones se le identifique todas sus órdenes de cultivos y de antibióticos. Luego a cada admisión se le va a marcar como su inicio de infección el tiempo en que la primera orden fue identificada. Luego los datos serán guardados en una vista provisional.

Esta da como resultado que, de los 704 pacientes diagnosticados con sepsis, a 670 se les puede identificar el inicio de infección. A partir de ahora se trabajará con estos 670 pacientes pues a los demás no se les va a poder identificar un tiempo de inicio de sepsis.

2.6.2 Identificación inicio de disfunción orgánica

Para la identificación del inicio de disfunción orgánica hará a través del indicador SOFA. Donde como ya visto, una variación de 2 punto en este indica el inicio de disfunción. Lo primero es

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

entonces identificar las mediciones de las variables que influyen en este indicador. Se va a hacer una vista provisional para las variables que influyen en el SOFA con sus varias mediciones para cada paciente. Se crearon vistas para las variables bilirrubina, creatinina, recuento de plaquetas, índice de Glasgow, índice pafi, presión arterial media y si el paciente estuvo o no conectado a ventilador. Desafortunadamente no se tiene suficientes datos sobre las órdenes de vasopresores a los pacientes, por lo que para este caso no se puede incluir el efecto de esta variable en el indicador SOFA calculado.

Ya teniendo esto se harán funciones de SQL que utilicen las mediciones de cada vista para calcular el puntaje que aporta cada variable al SOFA. Como se puede observar en la Figura 1.

Luego se va a hacer una función para calcular el SOFA total para un paciente a una hora determinada. Esta función toma la última medición de cada variable hasta la hora en que se va a calcular. Luego calcula el puntaje del SOFA según las mediciones obtenidas y los suma. También tiene en cuenta que el tiempo de la admisión sea consistente con el tiempo de medición de la variable, por lo que solo se van a considerar mediciones a partir de un día antes de ingreso a la UCI.

Por último, se va a realizar una función donde se tome para un paciente una ventana de tiempo de 72 horas alrededor del tiempo de inicio de infección. Desde 48 horas antes hasta 24 después. Se le va a calcular en cada hora el valor del indicador SOFA y cuando ocurra un cambio de 2 o más puntos en este, se marcará este tiempo como el inicio de la disfunción.

De aquí se le puede identificar el tiempo de inicio de disfunción para 314 registros de los 670. Para calcular el tiempo de inicio de la sepsis a cada uno de estos registros, se toma el máximo

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

de ambos tiempos de inicio, en caso de que no tenga tiempo de inicio de disfunción se toma el de inicio de infección.

Se define entonces que se va a trabajar con 670 admisiones con sepsis y con 3741 sin sepsis. Los tiempos de inicio de sepsis para los primeros 670 se calculan como se vio en esta admisión. Para los otros 3741 se toma un tiempo al azar durante su estancia en UCI.

2.7 RECOLECCIÓN DE DATOS Y ADAPTACIÓN DEL MODELO A DATOS DE LA INSTITUCIÓN DE SALUD LOCAL

Ya teniendo bien definido el tiempo de inicio de la sepsis y teniendo definido que se va a trabajar con las 670 admisiones anteriormente mencionadas se puede empezar a recolectar los atributos del modelo.

Con el tiempo de inicio de la sepsis se puede construir la ventana de tiempo en la cual se van a recolectar los atributos. Esta va a ser construida de la misma manera en que se construyó para la base de datos MIMIC, donde se definieron los parámetros t y $delta$ y se arma como se muestra en la Figura 8. Para este caso se tomaron ventanas de tiempo con $t=1$ y se varía el parámetro $delta$ entre $6-t$, $12-t$ y $24-t$. Se va a comparar el desempeño de los modelos en estas ventanas de tiempo.

Cada una de las variables se debe ubicar en la tabla de la base de datos que la contenga, se deben consultar sus mediciones y filtrar las que estén por fuera de la ventana de tiempo. Luego, se debe hacer un cruce entre las mediciones de esta variable y la tabla de las admisiones de los pacientes y por último hacer una agrupación y calcular los atributos del modelo. Nótese que muchas

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

de los datos estáticos son funciones agregadas como mínimos, máximos y promedios, véase **Anexo A**

Con la información que se tenía disponible en la base de datos, se pudo recolectar 34 de las 37 variables de entrada del modelo de datos estáticos para los pacientes. Véase nuevamente **Anexo A**. Sin embargo, estas contienen una gran cantidad de datos nulos. Esto la cantidad de nulos depende de la ventana de tiempo que se tenga (parámetro *delta*), donde una más grande evita que haya menos nulos. Inicialmente, se tiene que para $delta=6-t$, $12-t$ y $24-t$ el porcentaje de atributos nulos es de 39%, 35% y 31% respectivamente.

Para mitigar esto vamos a quedarnos con las instancias que tengan por lo menos siete (7) variables de estas no nulas. Esto da como resultado tres (3) nuevos conjuntos de entrenamiento para las ventanas de tiempo con $delta=6-t$, $12-t$ y $24-t$. Estos contienen 3512 pacientes sin sepsis y 457 pacientes con sepsis, 3579 pacientes sin sepsis y 457 pacientes con sepsis y 3649 pacientes con sepsis y 458 pacientes sin sepsis respectivamente. Realizando este filtrado los porcentaje de variables nulas en cada conjunto de datos son de 34%, 31% y 27% respectivamente. Donde se puede ver que hay una reducción considerable de las variables nulas y se asegura que cada instancia contenga datos para ser clasificada.

A continuación, se van a probar algunos de los modelos que ya se tenía para clasificar estas nuevas instancias, para ver si el modelo entrenados con los datos de MIMIC generaliza bien para estos.

En caso de que no se obtenga un desempeño satisfactorio, se utilizará y validará un nuevo modelo con los datos de la institución de salud local. Esto no necesariamente significa que el modelo inicial

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

no funcione, es posible que existan diferencias fundamentales entre los datos MIMIC y los de la institución de salud local, como por ejemplo la demografía de los pacientes, diferencias étnicas de los pacientes, diferencias de medición de las variables, inexistencia de variables importantes, etc.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

3. PRESENTACIÓN Y DISCUSIÓN DE RESULTADOS

Tabla 1. Resumen de resultados validación cruzada

Identificación de sepsis en MIMIC $t=0$, $\delta=4$ y 24				
Modelo	AUROC	Accuracy	Precision	Recall
kNN + función desface	0.664	0.615	0.609	0.641
kNN + DTW	0.822	0.785	0.864	0.675
Naive Bayes	0.775	0.71	0.73	0.71
Bosque Aleatorio	0.895	0.808	0.782	0.858
Xgboost	0.937	0.857	0.858	0.855
Multclasificador con Arbol de desición	0.855	0.855	0.866	0.831
Multclasificador con Máquina vector soporte	0.941	0.859	0.873	0.86
Multclasificador con Red neuronal multicapa	0.936	0.856	0.854	0.859
Multclasificador con Regresión logística	0.944	0.864	0.866	0.861
Predicción de sepsis en MIMIC $t=1$, $\delta=24-t$				
Multclasificador con Regresión logística	0.916	0.829	0.812	0.858
Xgboost	0.918	0.836	0.818	0.866
Predicción de sepsis en MIMIC $t=2$, $\delta=24-t$				
Xgboost	0.912	0.836	0.816	0.868
Predicción de sepsis en MIMIC $t=3$, $\delta=24-t$				
Xgboost	0.908	0.826	0.81	0.855

En la **¡Error! No se encuentra el origen de la referencia.** se puede observar un resumen de los modelos utilizados durante este trabajo y las métricas obtenidas por cada uno al realizar una validación cruzada. En la primera sección de la tabla se observan los modelos utilizados para la identificación de sepsis. Los modelos tipo kNN + función distancia utilizaron los datos de series de tiempo. Se puede ver que el desempeño de usar el algoritmo DTW supera de manera significativa en todas las métricas observadas a la *función desface*. En particular la métrica AUROC mejora de 0.664 a 0.822, que significa pasar de tener un desempeño mediocre a uno relativamente bueno. Debido a esto es evidente que la salida el modelo kNN con DTW, otorga más información sobre si

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

el paciente desarrolla sepsis o no por lo que éste va a ser utilizado más adelante en los métodos del multclasificador.

Más adelante se pueden ver los modelos que utilizaron con los datos estáticos, que son los datos clínicos del paciente, para la identificación de sepsis. Aquí se probaron modelos tipo *Naive Bayes*, Bosque Aleatorio y XGBoost. Como se puede ver el mejor desempeño fue obtenido por el modelo XGBoost con un AUROC de 0.937, que supera por un margen bastante grande al modelo *Naive Bayes* y por uno más pequeño pero evidenciable al Bosque Aleatorio. Por lo tanto, igual que el kNN con DTW, la salida del modelo XGBoost es la que más información aporta con estos datos sobre si el paciente desarrollo sepsis o no.

Luego se pueden observar los multclasificadores que utilizan los resultados del kNN con DTW y el XGBoost como entradas de un metaclasificador y los resultados que tienen a la hora de identificar la sepsis a la hora de inicio. El mejor modelo en cuanto a su métrica de AUROC es el que utiliza una regresión logística como metaclasificador. El modelo presenta una mejora en las métricas en comparación con únicamente utilizar el XGBoost. Este obtiene una métrica AUROC de 0.944 en la validación cruzada que es comparable con el estado del arte para este problema.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Tabla 2. Resumen de resultados conjunto de validación

Identificación de sepsis en MIMIC $t=0$, $\delta=4$ y 24				
Modelo	AUROC	Accuracy	Precision	Recall
kNN + DTW	0.83	0.847	0.635	0.702
Xgboost	0.93	0.836	0.585	0.852
Multclasificador con Regresión logística	0.941	0.843	0.597	0.861
Predicción de sepsis en MIMIC $t=1$, $\delta=24-t$				
Multclasificador con Regresión logística	0.906	0.798	0.497	0.857
Xgboost	0.912	0.809	0.539	0.876
Predicción de sepsis en MIMIC $t=2$, $\delta=24-t$				
Xgboost	0.909	0.804	0.531	0.884
Predicción de sepsis en MIMIC $t=3$, $\delta=24-t$				
Xgboost	0.905	0.792	0.514	0.866

Tabla 3. Resumen de resultados conjunto de validación más ajuste del umbral

Identificación de sepsis en MIMIC $t=0$, $\delta=4$ y 24				
Modelo	AUROC	Accuracy	Precision	Recall
Multclasificador con Regresión logística	0.941	0.894	0.755	0.758
Predicción de sepsis en MIMIC $t=1$, $\delta=24-t$				
Multclasificador con Regresión logística	0.906	0.859	0.642	0.672
Xgboost	0.912	0.852	0.667	0.64
Predicción de sepsis en MIMIC $t=2$, $\delta=24-t$				
Xgboost	0.909	0.851	0.666	0.634
Predicción de sepsis en MIMIC $t=3$, $\delta=24-t$				
Xgboost	0.905	0.844	0.655	0.6

Después de haber realizado una validación cruzada para estos modelos y determinado cual tiene un mejor desempeño y capacidad predictiva con estos datos, se va a entrenar el modelo multclasificador que utiliza la regresión logística con todos los datos del conjunto de entrenamiento y se va a evaluar según su desempeño en los datos de validación. Como se puede observar en la primera sección de la Tabla 2 este tiene una métrica de AUROC de 0.941, lo que significa que tiene un desempeño comparable al obtenido en la validación cruzada. Esta métrica también es mayor que utilizar el kNN con DTW o el XGBoost por separado para clasificar el conjunto de evaluación, La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

donde se puede observar nuevamente primera sección de la Tabla 2 que obtienen un AUROC de 0.83 y 0.93 respectivamente. Sin embargo, nótese que la métrica de *precision* que obtiene el modelo multclasificador es bastante más baja, esto se debe al gran desbalance que tienen los datos. En este caso se puede ajustar el umbral en que una instancia se considera como positiva para mejorar la métrica. Se ajustó el umbral de 0.5 a 0.9 con el objetivo de que las métricas de *precision* y *recall* tengan aproximadamente el mismo valor, las métricas resultantes se pueden ver primera sección de la Tabla 3. Nótese que este umbral puede ser ajustado según la métrica que se quiera priorizar a la hora de implementar el modelo, por ejemplo, en caso de que se pueda tener una tolerancia mayor por los falsos positivos (pacientes identificados incorrectamente como que desarrollaron sepsis) pero no por los falsos negativos (pacientes identificados incorrectamente como que no desarrollaron sepsis) se debe tener un umbral donde se le dé más importancia a la métrica de *recall*.

A continuación, se van a presentar los resultados de utilizar los modelos para predicción de sepsis. Para predecir con una (1) hora de anticipación, es decir con el parámetro $t=1$, se entrenó el modelo multclasificador utilizando el kNN con DTW para los datos de series de tiempo, el XGBoost para datos estáticos y una regresión logística como metaclasificador. Sin embargo, como ya fue mencionado en el capítulo de metodología, al incrementar el parámetro t disminuye el número de pacientes que tienen datos de series de tiempo, lo que implica que el modelo es entrenado con menos datos cosa que puede afectar su desempeño. Esto se ve reflejado en las métricas obtenidas que se muestran en la segunda sección de la Tabla 1, donde se obtiene un AUROC de 0.916 al hacer una validación cruzada, que, si bien sigue siendo un buen desempeño, tiene una disminución notable del mismo modelo en la sección anterior y es superado ligeramente por entrenar únicamente un XGBoost con el conjunto de entrenamiento completo. Como se puede

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

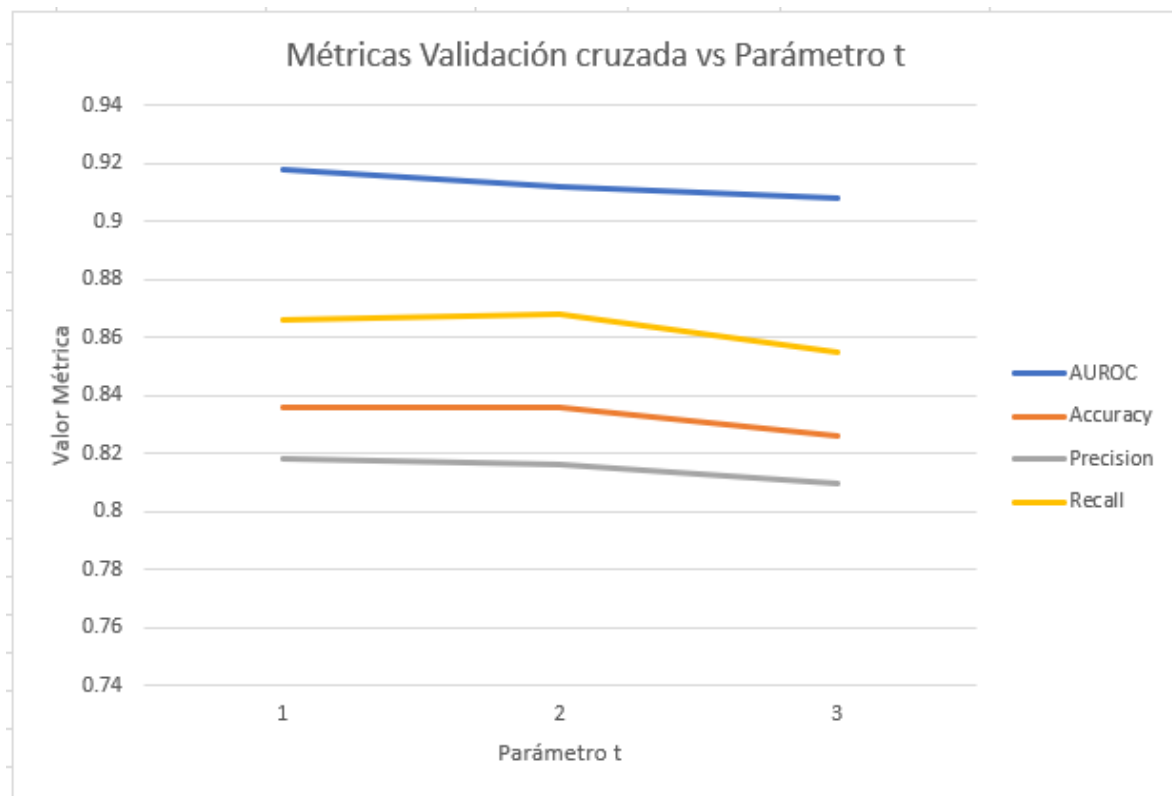
ver en la misma sección de la Tabla 1 este último tiene un AUROC de 0.918. Tanto el modelo multclasificador como el XGBoost siguen teniendo un desempeño similar a la hora de clasificar el conjunto de validación, con métricas AUROC de 0.906 y 0.912 respectivamente, presentadas en la segunda sección de la Tabla 2. En este caso el XGBoost también tiene un mejor desempeño.

Para el problema de predicción de sepsis también se puede ajustar el umbral a partir del cual se determina que una instancia desarrolla sepsis de acuerdo las métricas que se quieran priorizar. Las métricas resultantes se pueden ver en la Tabla 3.

A continuación, se presentan el desempeño de los modelos al predecir la sepsis con 1, 2 y 3 horas de anticipación. Se puede observar que a medida que se aumenta el parámetro t disminuye ligeramente el desempeño de modelo.

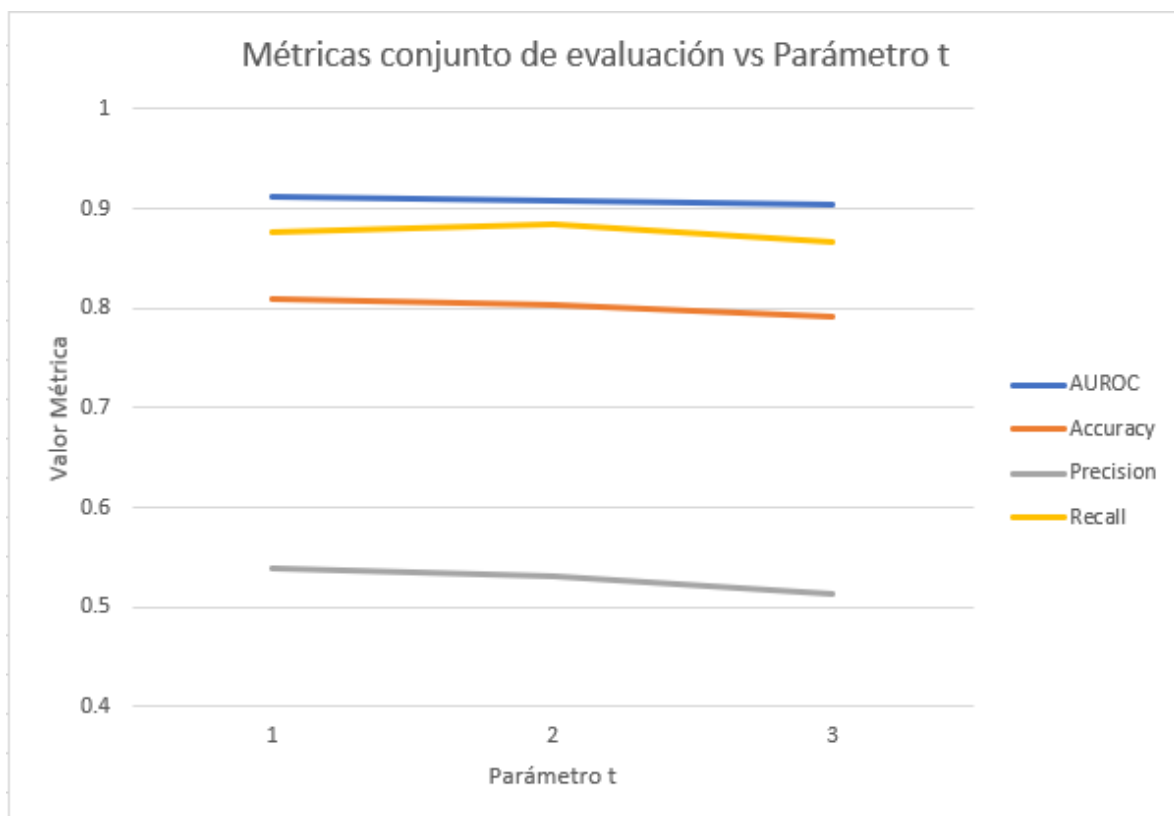
La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Ilustración 1. Comparación métricas validación cruzada según parametro t



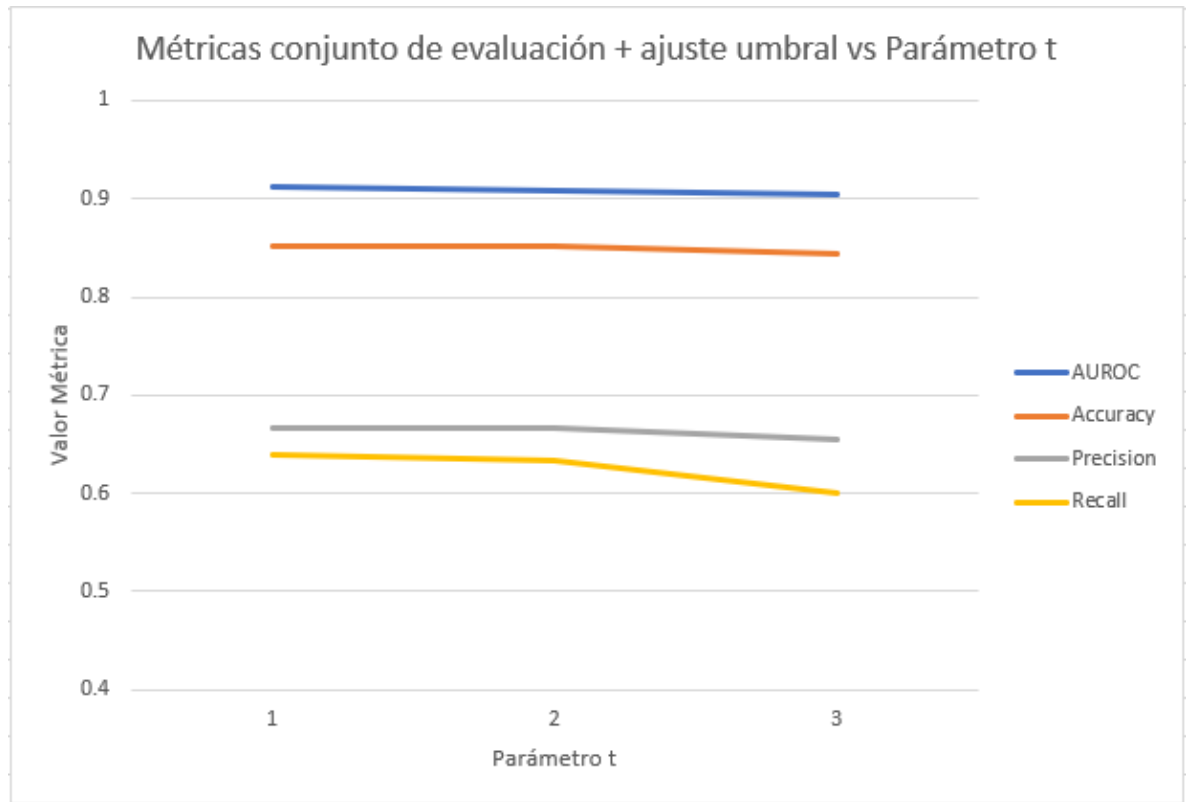
La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Ilustración 2. Comparación de métricas conjunto de validación según parámetro t



La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Ilustración 3. Comparación de métricas conjunto de evaluación con ajuste de umbral según parámetro t



En las gráficas anteriores se puede evidenciar esta ligera disminución en el desempeño del modelo.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Tabla 4. Resumen de resultados en la institución de salud local

Predicción de sepsis con datos de la institución de salud local $t=1$, $\delta=6-t$					
Modelo	Método de validación	AUROC	Accuracy	Precision	Recall
Xgboost (entrenado con datos de MIMIC)	Evaluación con datos de la institución	0.594	0.792	0.145	0.164
Xgboost	Validación cruzada	0.834	0.904	0.711	0.297
Predicción de sepsis con datos de la institución de salud local $t=1$, $\delta=12-t$					
Xgboost (entrenado con datos de MIMIC)	Evaluación con datos de la institución	0.59	0.78	0.125	0.157
Xgboost	Validación cruzada	0.848	0.907	0.698	0.324
Predicción de sepsis con datos de la institución de salud local $t=1$, $\delta=24-t$					
Xgboost (entrenado con datos de MIMIC)	Evaluación con datos de la institución	0.562	0.757	0.1	0.145
Xgboost	Validación cruzada	0.864	0.908	0.679	0.334

Por último, se presenta el desempeño de los modelos XGBoost para predecir sepsis con una (1) hora de anticipación utilizando los datos de la institución de salud local. En primer lugar se intentó hacer utilizar los modelos que se habían entrenado con los datos de MIMIC (el XGBoost con $t=1$), sin embargo como se puede ver en la Tabla 4 estos no dan en ningún caso buenos resultados para ningún parámetro δ . Donde la métrica AUROC no llega ni a 0.6. Por lo tanto, se realizó una validación cruzada con todo el conjunto de datos de pacientes de la institución variando el parámetro δ . A pesar de que esta vez se obtiene una métrica AUROC relativamente buena, se puede notar que se disminuye significativamente el desempeño del modelo al trabajar con estos datos con respecto a predecir sepsis en los datos de MIMIC. En particular, la métrica de recall es especialmente baja. Observando los datos se puede ver que esto se debe principalmente por la cantidad de datos nulos que se tienen en este conjunto, donde a pesar de hacer un filtrado de algunos datos nulos, todavía halla 34%, 31% y 27% de datos nulos respectivamente en cada conjunto. Esto indica que los datos no proporcionan la información necesaria para que los modelos funcionen adecuadamente.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

4. CONCLUSIONES Y CONSIDERACIONES FINALES

En este trabajo se propusieron modelos de *machine learning* con buen desempeño para la identificación temprana y predicción de sepsis que tienden el potencial para ser una pieza clave para abordar la problemática de sepsis en las unidades de cuidados intensivos. Para la identificación temprana de sepsis, con el modelo multclasificador se obtuvieron muy buenas métricas, resaltando el AUROC de 0.944.

Similarmente se pudo hacer una predicción de la sepsis con 1, 2 y 3 horas de anticipación con un AUROC de 0.918, 0.912, 0.908 respectivamente. Esto confirma el potencial que tienen los modelos de inteligencia artificial para abordar el problema de predicción de sepsis. Se puede ver que el desempeño de los modelos para este problema se ve afectado significativamente por la cantidad de datos que se tengan disponible, lo que muestra la importancia de una buena recolección de datos para este problema, en particular de datos de signos vitales, de laboratorio y algunos otros fisiológicos.

Se requiere seguir investigando los datos de la institución, aunque se pudo hacer pruebas efectivas con el modelo XGBoost, el desempeño de este va a estar limitado por la gran cantidad de datos nulos que se tienen. Esto indica que la recolección y el uso de los datos de pacientes junto a modelos de inteligencia artificial tiene un potencial grande para aplicaciones relacionadas con la salud de estos. Actualmente la mayor limitante para el desempeño de estos modelos es la cantidad de datos que se tienen disponibles en la base de datos de la institución de salud local.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Se sugiere continuar investigando la predicción de sepsis a partir de signos vitales, en este trabajo se trataron como series de tiempo y se utilizó un kNN junto al algoritmo DTW como función distancia para manejarlos. Se recomienda hacer una recolección de datos de signos vitales en la institución de salud local que tenga una resolución temporal mucho mayor que la de la base de datos de MIMIC e intentar obtener un mucho mejor desempeño con el kNN junto con DTW. Esto puede servir para tener mejores métricas en la clasificación de los pacientes de la institución local de salud.

También se puede experimentar con métodos redes neuronales recurrentes y redes neuronales convolucionales muestran usualmente buenos resultados trabajando con series de tiempo. Los modelos que utilizan estos datos han mostrado resultados inicial prometedores y si mejoran sus métricas podría tener aplicaciones prácticas interesantes. En particular, un modelo con estos datos solo necesita alimentarse del monitor de signos vitales del paciente y podría hacer predicciones en tiempo real fácilmente.

Finalmente, también puede ser interesante medir el desempeño de estos modelos en tiempo real, es decir sin que necesariamente generen alertas o afecten de algún modo al personal de la UCI. Se podría probar los modelos propuestos en tiempo real, guardando registro de cuando detecte un paciente con sepsis. Luego comparar los registros del modelo con los pacientes que desarrollaron sepsis. El objetivo sería entonces confirmar que las métricas del modelo se mantengan.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

REFERENCIAS

- Agarwal, A. K., Wadhwa, S., & Chandra, S. (1994). XGBoost: A Scalable Tree Boosting System. *The Journal of the Association of Physicians of India*, 42(8), 665.
- Barton, C., Chettipally, U., Zhou, Y., Jiang, Z., Lynn-Palevsky, A., Le, S., ... Das, R. (2019). Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. *Computers in Biology and Medicine*, 109(April), 79–84. <https://doi.org/10.1016/j.combiomed.2019.04.027>
- Cloud, P. A. (2018). A comparative analysis of sepsis identification methods in an electronic database. *Developing and Managing Engineering Procedures*, 46(4), 139–144. <https://doi.org/10.1016/b978-081551448-0.50008-0>
- Cortes, C., & Mohri, M. (2004). AUC optimization vs. Error rate minimization. *Advances in Neural Information Processing Systems*.
- Desautels, T., Calvert, J., Hoffman, J., Jay, M., Kerem, Y., Shieh, L., ... Das, R. (2016). Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. *JMIR Medical Informatics*, 4(3), e28. <https://doi.org/10.2196/medinform.5909>
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition*.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

- Guillermo Ortíz, Carmelo Dueñas, Ferney Rodríguez, Lena Barrera, Gisela de La Rosa, Rodolfo Dennis, Marcela Granados, Darío Londoño, Francisco Molina, F. J. (2014). *Epidemiología de la sepsis en unidades de cuidado intensivo en Colombia*. Retrieved from <https://revistabiomedica.org/index.php/biomedica/article/view/1439>
- Gyawali, B., Ramakrishna, K., & Dhamoon, A. S. (2019). Sepsis: The evolution in definition, pathophysiology, and management. *SAGE Open Medicine*, 7, 205031211983504. <https://doi.org/10.1177/2050312119835043>
- Lee, E. P. F., Lee, E. P. F., Lozeille, J., Soldán, P., Daire, S. E., Dyke, J. M., & Wright, T. G. (2001). An empirical study of the naive Bayes classifier. *Physical Chemistry Chemical Physics*, 3(22), 4863–4869. <https://doi.org/10.1039/b104835j>
- Mao, Q., Jay, M., Hoffman, J. L., Calvert, J., Barton, C., Shimabukuro, D., ... Das, R. (2018). Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. *BMJ Open*, 8(1), 1–11. <https://doi.org/10.1136/bmjopen-2017-017833>
- Marshall JC, Bosco L, Adhikari NK, Connolly B, Diaz JV, Dorman T, Fowler RA, Meyfroidt G, Nakagawa S, Pelosi P, Vincent JL, Vollman K, Z. J. (2017). *What is an intensive care unit? A report of the task force of the World Federation of Societies of Intensive and Critical Care Medicine*. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S0883944116302404>
- Michie, Donald. (1968). Memo Functions and Machine Learning. *Nature*, 218, 19–22.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

- Moor, M., Horn, M., Rieck, B., Roqueiro, D., & Borgwardt, K. (2019). *Early Recognition of Sepsis with Gaussian Process Temporal Convolutional Networks and Dynamic Time Warping*. Retrieved from <http://arxiv.org/abs/1902.01659>
- Nachimuthu, S. K., & Haug, P. J. (2012). Early detection of sepsis in the emergency department using Dynamic Bayesian Networks. *AMIA ... Annual Symposium Proceedings / AMIA Symposium. AMIA Symposium, 2012*, 653–662.
- Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *Journal of Educational Research*, 96(1), 3–14. <https://doi.org/10.1080/00220670209598786>
- Ranawana, R., & Palade, V. (2016). Multi-Classifer Systems: Review and a roadmap for developers. *International Journal of Hybrid Intelligent Systems*, 3(1), 35–61. <https://doi.org/10.3233/his-2006-3104>
- Shashikumar, S. P., Stanley, M. D., Sadiq, I., Li, Q., Holder, A., Clifford, G. D., & Nemati, S. (2017). Early sepsis detection in critical care patients using multiscale blood pressure and heart rate dynamics. *Journal of Electrocardiology*, 50(6), 739–743. <https://doi.org/10.1016/j.jelectrocard.2017.08.013>
- Singer, M., Deutschman, C. S., Seymour, C., Shankar-Hari, M., Annane, D., Bauer, M., ... Angus, D. C. (2016). The third international consensus definitions for sepsis and septic shock (sepsis-3). *JAMA - Journal of the American Medical Association*, 315(8), 801–810. <https://doi.org/10.1001/jama.2016.0287>

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Westphal, G. A., Koenig, Á., Filho, M. C., Feijó, J., de Oliveira, L. T., Nunes, F., ... Gonçalves, A. R. R. (2011). Reduced mortality after the implementation of a protocol for the early detection of severe sepsis. *Journal of Critical Care*, 26(1), 76–81.
<https://doi.org/10.1016/j.jcrc.2010.08.001>

Yasser Sakr, Ulrich Jaschinski, Xavier Wittebole, Tamas Szakmany, Jeffrey Lipman, Silvio A Namendys-Silva, Ignacio Martin-Loeches, Marc Leone, Mary-Nicoleta Lupu, J.-L. V. (2018). *Sepsis in Intensive Care Unit Patients: Worldwide Data From the Intensive Care over Nations Audit*.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Anexo A. Variables recolectadas de pacientes

variable	Descripción	Obtenida en MIMIC	Obtenida en la institución de salud local
edad	Edad del paciente en el momento que entro a UCI	SI	SI
peso	Peso del paciente	SI	SI
cancer	Si el paciente tuvo algún tipo de cáncer o no	SI	SI
diabetes	Si el paciente tuvo algún tipo diabetes o no	SI	SI
si_max	Shock index máximo que tuvo el paciente durante la ventana de tiempo	SI	SI
prueba2	presión sistólica * edad del paciente	SI	SI
elixhauser_hospital	indicador elixhauser del paciente	SI	NO
vent_predict	Si el paciente tuvo probabilidad de haber estado conectado a ventilador artificial	SI	NO

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

vaso_predict	Si es probable que al paciente le fueran administrados vasopresores	SI	NO
mingcs	Mínima medición del coeficiente de glasgow	SI	SI
sysbp_min	Presión arterial sistólica mínima que tuvo el paciente en la ventana de tiempo	SI	SI
sysbp_max	Presión arterial sistólica máxima que tuvo el paciente en la ventana de tiempo	SI	SI
resprate_min	Frecuencia respiratoria mínima que tuvo el paciente en la ventana de tiempo	SI	SI
resprate_max	Frecuencia respiratoria máxima que tuvo el paciente en la ventana de tiempo	SI	SI
diasbp_max	Presión arterial diastólica mínima que tuvo el paciente en la ventana de tiempo	SI	SI
diasbp_min	Presión arterial diastólica máxima que tuvo el paciente en la ventana de tiempo	SI	SI
meanbp_max	Presión arterial media mínima que tuvo el paciente en la ventana de tiempo	SI	SI

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

meanbp_min	Presión arterial media máxima que tuvo el paciente en la ventana de tiempo	SI	SI
spo2_min	Saturación de oxígeno mínima que tuvo el paciente en la ventana de tiempo	SI	SI
spo2_mean	Saturación de oxígeno media que tuvo el paciente en la ventana de tiempo	SI	SI
tempc_max	Temperatura máxima que tuvo el paciente en la ventana de tiempo	SI	SI
tempc_min	Temperatura mínima que tuvo el paciente en la ventana de tiempo	SI	SI
heartrate_max	Frecuencia cardiaca máxima que tuvo el paciente en la ventana de tiempo	SI	SI
heartrate_min	Frecuencia cardiaca mínima que tuvo el paciente en la ventana de tiempo	SI	SI
wbc_min	recuento de glóbulos blancos mínimo que tuvo el paciente en la ventana de tiempo	SI	SI
wbc_max	recuento de glóbulos blancos máximo que tuvo	SI	SI

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

	el paciente en la ventana de tiempo		
glucose_min	Glucosa mínima que tuvo el paciente en la ventana de tiempo	SI	SI
glucose_max	Glucosa máxima que tuvo el paciente en la ventana de tiempo	SI	SI
hematocrit_min	hematocritos mínimos que tuvo el paciente en la ventana de tiempo	SI	SI
hematocrit_max	hematocritos máximos que tuvo el paciente en la ventana de tiempo	SI	SI
hemoglobin_min	hemoglobina mínima que tuvo el paciente en la ventana de tiempo	SI	SI
hemoglobin_max	hemoglobina máxima que tuvo el paciente en la ventana de tiempo	SI	SI
platelet_min	recuento de plaquetas mínimo que tuvo el paciente en la ventana de tiempo	SI	SI
platelet_max	recuento de plaquetas máximo que tuvo el paciente en la ventana de tiempo	SI	SI

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

creatinine_min	Creatinina mínima que tuvo el paciente en la ventana de tiempo	SI	SI
creatinine_max	Glucosa máxima que tuvo el paciente en la ventana de tiempo	SI	SI

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.