



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

DISEÑO Y CONSTRUCCIÓN DE UN MODELO DE PREDICCIÓN DE MORTALIDAD EN
PACIENTES ADULTOS MAYORES HOSPITALIZADOS EN LA UNIDAD DE CUIDADOS
INTENSIVOS DEL HOSPITAL CLÍNICO DE LA UNIVERSIDAD DE CHILE

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERA CIVIL INDUSTRIAL

MARÍA BELÉN HARNECKER FONTECILLA

PROFESOR GUÍA:
JUAN VELÁSQUEZ SILVA

MIEMBROS DE LA COMISIÓN:
EDUARDO TOBAR ALMONACID
ROCÍO RUIZ MORENO

SANTIAGO DE CHILE
2017

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE: Ingeniera Civil Industrial
POR: María Belén Harnecker Fontecilla
FECHA: Agosto 2017
PROF. GUÍA: Juan Velásquez Silva

DISEÑO Y CONSTRUCCIÓN DE UN MODELO DE PREDICCIÓN DE MORTALIDAD EN
PACIENTES ADULTOS MAYORES HOSPITALIZADOS EN LA UNIDAD DE CUIDADOS
INTENSIVOS DEL HOSPITAL CLÍNICO DE LA UNIVERSIDAD DE CHILE

El presente trabajo de título tiene como objetivo diseñar y construir un modelo de predicción de mortalidad en adultos mayores admitidos en unidades de cuidados intensivos, utilizando información de salud específica de una cohorte de 403 pacientes admitidos en UCI del Hospital Clínico de la Universidad de Chile (HCUCH), almacenada bajo la forma de bases de datos disponibles en los sistemas de información del hospital.

A causa del envejecimiento de la población y del aumento de la esperanza de vida, las unidades de cuidados intensivos enfrentan nuevos desafíos: en el caso del HCUCH, casi la mitad de los pacientes que han pasado por este servicio en los últimos cinco años tienen más de 65 años. Adicionalmente, se ha demostrado que los pacientes mayores obtienen peores resultados después de un evento crítico comparado con el resto de la población adulta, lo cual se manifiesta tanto en el deterioro cognitivo y funcional, como en mayores tasas de mortalidad. Por otro lado, se identifica la oportunidad brindada por la masificación del uso de registros electrónicos de salud durante los últimos años, lo que da paso a la generación de grandes cantidades de información específica de cada paciente, almacenada bajo la forma de datos estructurados y no estructurados. Este conjunto problema-oportunidad, motiva al desarrollo de modelos que asignen a cada paciente un indicador de riesgo de mortalidad que, por un lado, apoye la toma de decisiones de los médicos, y por otro lado ayude a entender cuáles son los factores que explican un mayor riesgo y por lo tanto peores desenlaces clínicos.

Utilizando la metodología “Knowledge Discovery in Databases” (KDD), se construye un modelo de predicción de mortalidad intra-hospitalaria, mediante el entrenamiento de algoritmos de clasificación de data mining, empleando datos correspondientes a información demográfica y clínica de cada paciente, incluyendo diagnósticos, procedimientos, fármacos y exámenes de laboratorio, disponibles durante las primeras 24, 48 y 72 horas de hospitalización. Tanto el pre-procesamiento, la transformación y el modelamiento de la data fue realizado utilizando el lenguaje Python.

Los modelos construidos demuestran resultados satisfactorios: $AUC \approx 0.78-0.83$, $AUC \approx 0.82-0.85$ y $AUC \approx 0.81-0.82$ para modelos de 24, 48 y 72 horas respectivamente, desempeño comparable con los sistemas de scoring comúnmente utilizados en UCI (APACHE II y SOFA), teniendo la ventaja de ser modelos potencialmente útiles para desarrollar sistemas predictivos automáticos y dinámicos en el tiempo, además del gran impacto y aporte de ser construidos a partir de población chilena y específica de pacientes adultos mayores. Por otro lado la interpretación de los resultados resultan ser biológicamente plausibles, identificando factores de riesgo que pueden ser motivo de futuras investigaciones e iteraciones sobre el modelo propuesto. Se concluye que los resultados de este trabajo alientan seguir una línea de investigación al respecto.

A mi bisabuelo, Reinaldo Harnecker V.

Agradecimientos

En primer lugar, quiero agradecer a mis papás, Reinaldo y Josefina. Agradezco infinitamente el apoyo y la confianza sin condiciones que me han entregado a lo largo de mi vida, en todas las decisiones y caminos que he tomado. Gracias a mi papá por toda esa energía, por todas las ganas y la insistencia para hacerme siempre querer llegar más lejos, de querer poner siempre lo mejor de mí. Gracias a mi mamá por su amor tan grande, por sus consejos, por su humor y por su fortaleza que tanto admiro.

A la Ema, por su entrega y amor incondicional.

A mi abuelo Víctor, por creer en mí y alentarme a seguir este camino.

A mis hermanos, Pedro, Santiago y Reinaldo, por existir.

A mis profesores, Juan y Eduardo, por su dedicación e interés, por guiarme y por darme la oportunidad de tomar este desafío. También quiero agradecer al centro por la acogida y por el apoyo, a la Rocío y a Felipe por su tiempo, su guía y sus consejos.

A Felipe Salech y todas las personas del Hospital Clínico de la Universidad de Chile que me ayudaron en este proceso, por abrirme las puertas, por la ayuda y por el tiempo dedicado, sin los cuales no habría sido posible llevar a cabo esta investigación.

Finalmente quiero agradecer mi paso por la facultad de ingeniería, quien me haya conocido el primer día de universidad sabrá todo lo que me han hecho crecer estos siete años. Gracias por todo lo aprendido y lo vivido, y gracias por darme la oportunidad de conocer a las personas que me han acompañado en esta etapa, mención especial a mis mejores amigas, la Antonia y la Pauli, a quienes agradezco por la linda amistad que hemos construido, por su compañía que ha sido fundamental en los momentos más felices y en los más difíciles de estos tiempos universitarios, sin duda son lo más valioso que me llevo.

Tabla de Contenido

Índice de Tablas	vii
Índice de Figuras	viii
1. Introducción	1
1.1. Antecedentes Generales	1
1.1.1. WIC	1
1.1.2. Hospital Clínico de la Universidad de Chile	2
1.2. Justificación	3
1.2.1. Unidades de Cuidados Intensivos	3
1.2.2. Adultos Mayores Hospitalizados en Unidades de Cuidados Intensivos	4
1.2.3. Herramientas de Apoyo a la Decisión	7
1.2.4. Propuesta de solución	8
1.3. Hipótesis de Investigación	8
1.4. Objetivos	8
1.4.1. Objetivo General	8
1.4.2. Objetivos Específicos	8
1.5. Metodología	9
1.6. Resultados Esperados	10
1.7. Alcances	10
1.8. Estructura del Informe	11
2. Marco Teórico	13
2.1. Data Mining	13
2.1.1. KDD y Data Mining	13
2.1.2. Machine Learning	15
2.1.3. Aprendizaje Supervisado y No Supervisado	15
2.2. Missing Data	15
2.2.1. Tipos de Valores Faltantes	15
2.2.2. Manejo de Datos Faltantes	16
2.3. Selección de Atributos	17
2.3.1. F-Score	18
2.3.2. Correlación de Pearson	19
2.4. Métodos de Clasificación	20
2.4.1. Regresión Logística	20
2.4.2. Naïve Bayes	21
2.4.3. Support Vector Machines(SVM)	21

2.4.4. Redes Neuronales Artificiales	22
2.5. Validación de Modelos	23
2.5.1. Entrenamiento y Prueba del Modelo	24
2.5.2. Tipos de Error y Medidas de Evaluación	26
2.5.3. Curva Receiver Operating Characteristics (ROC)	27
3. Modelos de Predicción de Mortalidad	29
4. Diseño del Modelo	32
4.1. Selección y Extracción de Datos	32
4.1.1. Ficha Clínica Electrónica	32
4.1.2. Farmacia	33
4.1.3. Laboratorio	34
4.1.4. Base de Datos Preliminar	34
4.1.5. Diseño del cohorte	35
4.1.6. Estadística Descriptiva	37
4.2. Selección y Transformación de Variables	41
4.2.1. Transformación Variables Diagnóstico Principal	41
4.2.2. Transformación Variables Diagnósticos Secundarios	43
4.2.3. Transformación Variables Procedimientos	45
4.2.4. Transformación Variables Farmacia	45
4.2.5. Transformación Variables Exámenes de Laboratorio	46
4.3. Manejo de Datos Faltantes	47
5. Modelamiento	51
5.1. Selección de Variables	51
5.2. Modelo 24 horas	54
5.2.1. Selección de Variables Modelo 24 horas	54
5.2.2. Resultados Entrenamiento Modelo 24 horas	57
5.2.3. Interpretación Modelo 24 horas	59
5.3. Modelo 48 horas	63
5.3.1. Selección de Variables Modelo 48 horas	63
5.3.2. Resultados Entrenamiento Modelo 48 horas	67
5.3.3. Interpretación Modelo 48 horas	69
5.4. Modelo 72 horas	72
5.4.1. Selección de Variables Modelo 72 horas	72
5.4.2. Resultados Entrenamiento Modelo 72 horas	75
5.4.3. Interpretación Modelo 72 horas	77
5.5. Discusión	81
5.5.1. Análisis y Comparación de Resultados	81
5.5.2. Opinión Experta	83
6. Conclusiones y Trabajo Futuro	85
6.1. Conclusiones	85
6.2. Trabajo Futuro	87
7. Bibliografía	89

A. Resultados Análisis Puntaje F	96
A.1. Variables y puntaje F para corte p-valor=0.3 para modelo 24 horas	96
A.2. Variables y puntaje F para corte p-valor=0.3 para modelo 48 horas	99
A.3. Variables y puntaje F para corte p-valor=0.25 para modelo 72 horas	102
B. Coeficientes Logit	106
B.1. Variables y coeficiente β para modelo 24 horas	106
B.2. Variables y coeficiente β para modelo 48 horas	109
B.3. Variables y coeficiente β para modelo 72 horas	112

Índice de Tablas

3.1. Sistemas Scoring Clásicos	30
3.2. Modelos predictivos machine learning	31
4.1. Descripción de atributos set de datos ficha clínica electrónica	33
4.2. Descripción variables exámenes de laboratorio	34
4.3. Resumen variables extracción inicial	35
4.4. Clasificación de comorbilidades de Elixhauser CIE-10	43
4.5. Clasificación de Medicamentos	46
4.6. Variables derivadas resultados laboratorio	47
4.7. Portcentaje de complitud variables de laboratorio por configuración de predicción	47
4.8. Variables de rango laboratorio	48
5.1. Resumen variables transformadas	52
5.2. Puntaje F modelo 24 horas	55
5.3. Resumen resultados promedio AUC de validación cruzada para cuatro casificado- res, según cada umbral de corte para base de datos 24 horas	56
5.4. Resumen resultados entrenamiento de modelo de 24 horas para cuatro clasifica- dores	57
5.5. Puntaje F modelo 48 horas	65
5.6. Resumen resultados promedio AUC de validación cruzada para cuatro casificado- res, según cada umbral de corte para base de datos 48 horas	66
5.7. Resumen resultados entrenamiento de modelo de 48 horas para cuatro clasifica- dores	67
5.8. Puntaje F modelo 72 horas	73
5.9. Resumen resultados promedio AUC de validación cruzada para cuatro casificado- res, según cada umbral de corte para base de datos 72 horas	74
5.10. Resumen resultados entrenamiento de modelo de 72 horas para cuatro clasifica- dores	75
A.1. Puntaje F variables relevantes, modelo 24 horas	96
A.2. Puntaje F variables relevantes, modelo 48 horas	99
A.3. Puntaje F variables relevantes, modelo 72 horas	102
B.1. Variables y coeficiente β para modelo 24 horas	106
B.2. Variables y coeficiente β para modelo 48 horas	109
B.3. Variables y coeficiente β para modelo 72 horas	112

Índice de Figuras

2.1. Proceso KDD	14
2.2. <i>Scatterplot de data set de dos clases, con hiperplanos de máxima distancia encontrados mediante algoritmo SVM con núcleo linear. Las instancias que coinciden en los hiperplanos de máxima distancia (puntos rojos) se llaman vectores de soporte</i>	22
2.3. Estructura MLP	23
2.4. <i>Esquema de división del conjunto de datos para validación cruzada de 4 folds</i>	25
2.5. <i>Validación cruzada "Leave one out"(LOOCV)</i>	25
2.6. <i>Estructura de una matriz de confusión para problema de clasificación binario</i>	26
2.7. <i>Espacio ROC</i>	28
4.1. Cantidad de observaciones disponibles por configuración de predicción	37
4.2. Histogramas variables demográficas y descriptores generales	38
4.3. Histograma variables laboratorio	39
4.4. Histograma variables laboratorio	40
4.5. Jerarquía CIE 10 representada por cuatro niveles	41
4.6. Distribución Diagnóstico-Paciente Nivel 4 Jerarquía CIE-10	42
4.7. Distribución Diagnóstico-Paciente Nivel 3 Jerarquía CIE-10	42
5.1. Mapa de calor covariables modelo 24 horas	54
5.2. Puntaje F modelo 24 horas	55
5.3. Gráfico promedio AUC de validación cruzada para cuatro clasificadores, según cada umbral de corte para base de datos 24 horas	56
5.4. Comparación curva ROC promedio para cuatro clasificadores en modelo 24 horas	58
5.5. Curva ROC para cuatro clasificadores en modelo 24 horas. La curva azul opaca corresponde al promedio de cada curva generada por las iteraciones de la validación cruzada, las cuales se muestran en azul claro	59
5.6. Coeficientes Logit modelo 24 horas	60
5.7. Puntaje F modelo 48 horas	64
5.8. Gráfico promedio AUC de validación cruzada para cuatro clasificadores, según cada umbral de corte para base de datos 48 horas	66
5.9. Curva ROC para cuatro clasificadores en modelo 48 horas. La curva azul opaca corresponde al promedio de cada curva generada por las iteraciones de la validación cruzada, las cuales se muestran en azul claro	68
5.10. Comparación curva ROC promedio para cuatro clasificadores en modelo 48 horas	69
5.11. Coeficientes Logit para 20 variables más relevantes, modelo 48 horas	70
5.12. Puntaje F modelo 72 horas	73

5.13. Gráfico promedio AUC de validación cruzada para cuatro clasificadores, según cada umbral de corte para base de datos 72 horas	74
5.14. Curva ROC para cuatro clasificadores en modelo 72 horas. La curva azul opaca corresponde al promedio de cada curva generada por las iteraciones de la validación cruzada, las cuales se muestran en azul claro	76
5.15. Comparación curva ROC promedio para cuatro clasificadores en modelo 72 horas	77
5.16. Coeficientes Logit modelo 72 horas	78
5.17. Comparación AUC	81

Capítulo 1

Introducción

1.1. Antecedentes Generales

El presente trabajo de título se desarrolla en el marco del Web Intelligence Centre de la Universidad de Chile (WIC), como un proyecto en conjunto entre la Facultad de Ciencias Físicas y Matemáticas y la Facultad de Medicina. Este nace desde la iniciativa del equipo docente del Hospital Clínico de la Universidad de Chile (HCUCH), compuesto por Eduardo Tobar, director académico de la Facultad de Medicina e Intensivista en HCUCH, y por Felipe Salech, becado del programa de formación de especialidad en geriatría HCUCH, planteando un problema del ámbito de la medicina que puede ser abarcado desde la ingeniería, específicamente desde la disciplina del análisis de datos.

El Dr. Eduardo Tobar ha destacado como investigador, enfocando su trabajo en el área de sedación, analgesia y delirium¹ en pacientes críticos. El Dr. Felipe Salech ha destacado en investigación en el área del envejecimiento. Anteriormente ambos participaron en iniciativas interdisciplinarias junto al WIC, en el marco de DELIRIUM, proyecto Fondef cuyo objetivo es el desarrollo de una aplicación portátil para implementar medidas de prevención de delirium en adultos mayores hospitalizados, que, al igual que el presente trabajo, combina intereses de especialidades de la medicina intensiva con geriatría, ofreciendo nuevas oportunidades de generar soluciones integradoras y colaborativas, aprovechando las capacidades y recursos que existen en la Facultad de Ciencias Físicas y Matemáticas

1.1.1. WIC

El Web Intelligence Centre es un centro de investigación que nace hace aproximadamente 10 años bajo la iniciativa del profesor Juan Velásquez, en el marco del departamento de Ingeniería Civil Industrial de la Universidad de Chile.

¹Condición de alteración del funcionamiento cerebral que se produce más frecuentemente en los adultos mayores que se hospitalizan y que se debe tanto a factores predisponentes, como también a factores precipitantes que tienen relación con la propia hospitalización

El WIC es miembro del Web Intelligence Consortium, una organización internacional sin fines de lucro, dedicada a la investigación científica e industrial en el campo de la Inteligencia Web. La misión, visión y objetivos del WIC se detallan en su página web²:

- Misión:
“Desarrollar investigación de frontera en el campo de Tecnologías de Información creando nuevas soluciones para abordar problemas complejos de ingeniería utilizando herramientas basadas en la Web de las Cosas”.
- Visión
“Ser un líder a nivel internacional en la investigación de tecnologías de información y comunicaciones aplicadas a la resolución de problemas del mundo real”.
- Objetivos
 - Publicar en las principales revistas, conferencias y editoriales relacionadas con Web Intelligence.
 - Proveer un servicio profesional, excelente y rápido para todos nuestros clientes.
 - Dictar cursos de orientación práctica acerca de las Tecnologías de Información y su aplicación en los negocios.

1.1.2. Hospital Clínico de la Universidad de Chile

El Hospital Clínico Universidad de Chile (HCUCH), también conocido como José Joaquín Aguirre o JJ Aguirre, es el principal campo clínico-docente de la Universidad de Chile y presta atención clínica ambulatoria, hospitalaria y de urgencia pediátrica, adulta y maternidad, destacándose por sus intervenciones de alta complejidad. Este conforma una Red Clínica junto a la Clínica Quilín, con una superficie total de 65.000, 3.307 funcionarios y 541 camas.

Como prestadora de salud, el HCUCH es una institución de carácter mixto: al ser parte de la Universidad, el HCUCH es una institución pública, pero a la vez ofrece prestaciones privadas. Se estima que cerca del 70% de las prestaciones son privadas, mientras que el restante 30% corresponde a prestaciones FONASA.

Los servicios ofrecidos por el HCUCH se agrupan en dos líneas:

- Servicios asistenciales de Salud: El hospital presta atención de tres tipos, Atención Ambulatoria, Atención Cerrada y Atención de Urgencia. La atención ambulatoria corresponde a las consultas de especialidad, procedimientos y cirugías ambulatorias. La atención cerrada corresponde a los servicios de hospitalización, ofreciendo servicios de adulto, pediátrico y maternidad, teniendo una capacidad total de 541 camas, de las cuales 55 son de cuidados intensivos y de intermedios, 60 camas de obstetricia y ginecología, 67 para pensionados y 456 camas básicas.
- Servicios Académicos: Siendo el principal campo clínico de la Universidad de Chile, el HCUCH entrega formación de pregrado y post grado de la carrera de medicina, y de otras carreras de la salud, como enfermería, kinesiología y terapia ocupacional. Además del área

²<http://wic.uchile.cl/>

formativa, el HCUCH realiza esfuerzos en crear y fortalecer la investigación clínica a través de la subdirección de investigación, siendo esta parte importante de su misión, en línea con su carácter de hospital universitario.

1.2. Justificación

1.2.1. Unidades de Cuidados Intensivos

La Medicina Intensiva se define como aquella área de la medicina que se ocupa de los pacientes con alteraciones fisiopatológicas que hayan alcanzado un nivel de severidad tal que representen una amenaza actual o potencial para su vida, y al mismo tiempo, sean susceptibles de recuperación.

Las Unidades de Cuidados Intensivos (UCI), son servicios clínicos centrales donde se realiza la labor asistencial de la especialidad. Estos operan estrechamente vinculados con el resto de los servicios hospitalarios, abarcando a una amplia gama de pacientes y diagnósticos clínicos, los cuales tienen en común el carácter crítico y potencialmente recuperable. Para ello estos servicios disponen de características especiales de diseño, equipamiento y personal.

- Equipos de ventilación mecánica (VM) para asistir a la respiración mediante un tubo endotraqueal o una traquetomía
- Equipos de hemodiálisis para terapias de reemplazo renal
- Equipo de monitorización cardiovascular avanzados
- Vías endovenosas para infusiones farmacológicas o para nutrición parenteral total
- Tubos nasogástricos, bombas de succión, drenajes y catéteres.
- Una amplia gama de fármacos incluyendo fármacos vaso activos, sedantes, antibióticos de amplio espectro y analgésicos

Los pacientes admitidos en UCI pueden agruparse en dos tipos: los pacientes médicos, aquellos que son admitidos consecuencia de enfermedades o lesiones críticas, y pacientes quirúrgicos; aquellos que son admitidos para la supervisión intensiva después de una cirugía mayor, los cuales a su vez pueden caer en dos categorías: pacientes quirúrgicos de urgencia, y pacientes programados. A pacientes médicos y pacientes quirúrgicos de urgencia generalmente se les asocia a peores resultados y mayor riesgo de mortalidad. La Unidad de Cuidados Intensivos del HCUCH se define como una UCI médico-quirúrgica de alta complejidad. Respecto a la admisión, entre un 20% y un 30% de los pacientes ingresan por urgencia, y entre un 70% y un 80% ingresan por admisión. Tiene una capacidad de 12 camas, y atiende fundamentalmente a adultos, siendo de 64 años el promedio de edad de los pacientes admitidos.

1.2.2. Adultos Mayores Hospitalizados en Unidades de Cuidados Intensivos

Ante el cambio demográfico que se presenta tanto en Chile como en el resto del mundo, las Unidades de Cuidados Intensivos enfrentan un nuevo desafío. En Chile, para la población de 60 años o más, se espera un crecimiento desde 13,7% en 2013 a un 17,3% en 2020 [15]. Sumado al envejecimiento de la población, se ha desarrollado un fuerte aumento de la esperanza de vida, desde 77,1 años en 2002 a un proyectado de 79,73 años en 2020.

Por otro lado, la Encuesta Nacional de Salud del año 2010 muestra que un 30% del total de las hospitalizaciones corresponden a personas mayores, siendo superior al 50% en servicios de Medicina Interna y de Paciente Crítico, con un creciente número de adultos mayores de 80 años[14]. Según estudios externos, los pacientes mayores de 80 años en unidades críticas representan entre un 10% y un 20% de los ingresos [48].

En términos de impacto económico, según el estudio “Impacto del envejecimiento en el gasto en la salud: Chile 2002 - 2020” generado por la Superintendencia de Salud, que las hospitalizaciones de adultos mayores al año 2002, generó un 50% del gasto en salud en dicha población y que para el año 2020 será 2,5 veces mayor.[16]

Con todo lo anterior, se ha comprobado que efectivamente la demanda por servicios de salud por parte de adultos mayores ha ido en aumento en los últimos años, y como consecuencia, las unidades de pacientes críticos enfrentan una creciente demanda por servicios de cuidados intensivos, y con esto, deben hacerse cargo de las características y problemas específicos de salud que representan a este grupo demográfico.

Por otro lado, ante este aumento acelerado de la demanda, diversos estudios han demostrado que el beneficio de los cuidados intensivos en adultos mayores puede ser controversial, lo cual se evidencia a través de los pobres resultados obtenidos tras este tipo de tratamientos, representados por una alta mortalidad intra-hospitalaria y en la declinación de funciones físicas y cognitivas post UCI.

Declinación Funcional:

Estudios internacionales demuestran una alta prevalencia de disfunciones físicas en adultos mayores sobrevivientes a una enfermedad crítica, presente entre un 33% y un 58% de los pacientes en seguimientos de 3 meses, y entre un 12% y un 97% en seguimientos de 6 meses. Por otro lado, estudios que utilizan funciones físicas base de los pacientes como variable de comparación, indican que entre un 10% y un 63% empeoran su condición inicial o reportan nuevas discapacidades. En cuanto a estudios realizados en torno al decaimiento de funciones cognitivas, indican prevalencia de demencia senil entre un 12% y un 18% de los pacientes después de padecer la enfermedad crítica.[10]

Numerosas investigaciones que estudian distintas poblaciones, reportan distintos resultados respecto de la declinación funcional de los pacientes que padecen de enfermedades críticas, sin embargo, en todos los casos los porcentajes son significativamente mayores comparados con otros grupos etarios y con adultos mayores hospitalizados por diagnósticos menos graves. Lo importante es destacar que estos desenlaces significan a largo plazo una peor calidad de vida para el paciente, mayor dependencia, mayor vulnerabilidad a padecer otras enfermedades, y

una carga emocional y financiera tanto para ellos mismos, como para sus familiares y para la sociedad.[21]

Mortalidad:

Estudios observacionales indican que la mortalidad en UCI en pacientes adultos mayores representa entre 12% y un 15%. La mortalidad al alta del hospital se estima alrededor de un 25%, y, luego de esto, se estima que entre un 40% y 50% muere después de 1 año. [48]

Una publicación realizada en Chile [67], que estudia las características de pacientes ventilados mecánicamente en 19 UCIs del país, reporta que la mortalidad en pacientes adultos mayores de 60 años ventilados mecánicamente es de un 43,5%, significativamente mayor que en el caso de pacientes menores de 60 años, la cual es de un 26,4%. Este hallazgo es de relevancia, ya que se estima que alrededor de un tercio de los pacientes ingresados en las UCIs participantes fue sometido a ventilación mecánica. Otro hallazgo a destacar, fue que la mortalidad del grupo de estudio es similar a la descrita para Latinoamérica y Europa, y que la mortalidad hospitalaria concuerda con la descrita en España. Dado lo anterior, se revisó la investigación de García Lizana et al. realizado en España el año 2001 [25], la cual estudia el pronóstico de 313 pacientes ≥ 65 años ingresados en la UCI. Encontraron una mortalidad del 30% en la UCI, 10% en el hospital y 11% tras el alta. La mortalidad al año fue del 51%. De los supervivientes, el 58,6% tenían buen estado de salud al año de seguimiento y eran independientes, el 24,3% tenían algún grado de discapacidad y el 17% eran dependientes. Se destaca que los resultados son similares a los estudios mencionados en el párrafo anterior.

Causas:

Además de la causal de hospitalización y de las características específicas del diagnóstico clínico, varias son las causas por las cuales los pacientes más ancianos tienen resultados más pobres en comparación al resto de los pacientes adultos en unidades de cuidados intensivos, algunas de estas son:

1. Base Funcional: Debido al proceso natural de envejecimiento, los adultos mayores presentan una disminución de reservas físicas y cognitivas, lo que implica que están más expuestos a padecer de enfermedades más complejas y a ser sometidos a procedimientos agresivos de soporte vital. Esto ocurre debido a la acumulación de daño celular, que conduce a disfunciones de los sistemas y órganos, lo que los expone a un mayor riesgo a las enfermedades críticas, a la dependencia y a la muerte. Por otro lado, la disminución funciones reparadoras (mecanismos homeostáticos), implica una recuperación más lenta después de una enfermedad crítica, una menor resistencia a tratamientos más complejos e invasivos, y muchas veces, la pérdida desproporcionada de su funcionalidad basal.
2. Inmovilidad: Debido a los cambios naturales del envejecimiento, las personas mayores presentan una disminución significativa de masa muscular y de fuerza (sarcopenia³), además de una disminución de su capacidad aeróbica. Esto, sumado a la situación de inmovilidad o reposo prolongado que es común en cuidado críticos, hacen que el paciente esté expuesto a una mayor pérdida de musculatura y fuerza, al discondicionamiento físico, lo que desencadena una declinación en su funcionalidad física y, por lo tanto, a una mayor

³Sarcopenia: pérdida degenerativa de masa muscular y fuerza al envejecer o al llevar una vida sedentaria

vulnerabilidad y dependencia.

3. Delirium: El paciente crítico puede experimentar delirium, consecuencia de la interacción de varios factores: el mismo diagnóstico, los fármacos delirógenos, la deshidratación, respuestas inflamatorias y daños metabólicos, entre otros. Se ha reportado que el delirium se presenta en el 80% de todos los pacientes ventilados mecánicamente [10]. Además, existe evidencia que asocia la presencia de delirium y secuelas cognitivas de largo plazo, más aún en pacientes que presentan demencia o disminución de funciones cognitivas previa a la hospitalización.).
4. Polifarmacia: Producto de la comorbilidad⁴ común en pacientes críticos AM, estos están más expuestos a interacciones farmacológicas. Esto puede afectar hasta un 15% de quienes ingresan [34].

Además de estas, existen muchas otras causas que contribuyen a explicar los malos resultados de los pacientes críticos adultos mayores, tales como la dependencia, la desnutrición, deshidratación, comorbilidad, cáncer activo, resucitaciones cardio - pulmonares, etc.

A modo de resumen, se destaca que los peores resultados que sufren pacientes críticos mayores, reflejados en la alta tasa de mortalidad y la significativa incidencia de disfunciones físicas y cognitivas consecuentes a enfermedades críticas, se debe a la interacción de tres grandes efectos:

1. El estado de salud basal de pacientes mayores y su vulnerabilidad ante enfermedades críticas
2. Los efectos de la enfermedad crítica subyacente
3. Los procedimientos y tratamientos impartidos durante la estadía en UCI

Dado lo anterior, y dado que no es posible combatir los efectos del proceso natural de envejecimiento, para prevenir o disminuir los malos resultados en este grupo demográfico, la medicina intensiva pone el foco en identificar cuáles son los pacientes más riesgosos, cuáles son los factores que se asocian con peores resultados, y así poder discernir cuáles pacientes se beneficiarían del tratamiento UCI y cuáles no, pudiendo optar por cuidados intermedios o tratamientos paliativos⁵.

Desde aquí deriva la necesidad de desarrollar una urgente definición de cuáles son los cuidados, tratamientos y procedimientos óptimos para atender pacientes adultos mayores con enfermedades críticas, tanto desde el punto de vista de la ética como de la evidencia objetiva. La elección y uso de los tratamientos mencionados, se manifiesta en la toma de decisiones que lleva a cabo el médico tratante a lo largo del período de hospitalización del paciente crítico.

⁴Comorbilidad: La presencia de uno o más diagnósticos además del diagnóstico principal

⁵Los Cuidados Paliativos son un modo de abordar la enfermedad crítica que pretende mejorar la calidad de vida tanto de los pacientes como de sus familias, mediante la prevención y el alivio del sufrimiento a través de un diagnóstico precoz, una evaluación adecuada y el oportuno tratamiento del dolor y de otros problemas tanto físicos como psicosociales y espirituales (definición OMS)

1.2.3. Herramientas de Apoyo a la Decisión

Tradicionalmente, las decisiones clínicas han recaído en la opinión basada en la experiencia individual del médico o equipo tratante. Durante las últimas décadas la medicina ha sufrido cambios drásticos, pasando de enfatizar en el diagnóstico y la observación del paciente, a poner en primer lugar los asuntos terapéuticos, específicamente en desarrollar tratamientos multidisciplinarios muy complejos dirigidos a pacientes con enfermedades más graves, los cuales se concentran en su mayoría en las UCIs.

Desde este nuevo paradigma de la medicina intensiva, sumado a la inversión en tecnologías, al alto costo de los recursos, y a la creciente demanda que requiere este tipo de cuidados, surge la necesidad de tener una evidencia más objetiva de los beneficios de tratamientos intensivos a la hora de tomar decisiones. Una forma de hacer esto es estimar la probabilidad de ocurrencia de ciertos desenlaces clínicos [41], es decir, poder clasificar a cada paciente con un indicador de riesgo, aportando en mejorar el entendimiento clínico sobre las variables que afectan en peores resultados.

En este contexto nacen las reglas de decisión y los sistemas de scoring. En 1981, Knaus, Zimmerman y Wagner desarrollan el Acute Physiology and Chronic Health Evaluation o APACHE, un sistema de clasificación de gravedad para pacientes críticos, que nace bajo la hipótesis de que la gravedad de la enfermedad crítica puede ser medida cuantificando el grado de anomalía de un conjunto de variables fisiológicas del paciente [41]. En 1985 nace su segunda versión, APACHE II, el cual es altamente utilizado hasta el día de hoy en UCIs de todo el mundo.

Posteriormente se han desarrollado otros sistemas de scoring para unidades de cuidados intensivos que cumplen con objetivos similares, actualmente en el HCUCH y en general en los establecimientos del sistema de salud chileno, los sistemas de scoring que se usan con mayor frecuencia son APACHE II y SOFA. Ambos se describen detalladamente en el capítulo 3 del presente trabajo.

Por otro lado, se observa que ninguno de los sistemas scoring que se utilizan actualmente se adaptan a características específicas de los adultos mayores. En consecuencia, se concluye que no existe un indicador validado que de luces acerca de cuáles pacientes adultos mayores se beneficiaría de la UCI, lo que hace que sea muy variable el criterio de admisión a unidades críticas según el médico y/o la institución tratante [48]. Además, estos sistemas de scoring tienen una serie de otras falencias:

- Los métodos de cálculo son puramente estadísticos, por lo cual no indican la probabilidad de supervivencia real individual de cada paciente.
- No se ajustan igual para todas las poblaciones
- Utilizan una cantidad limitada de variables
- Excluyen factores potencialmente importantes para la predicción de desenlaces, que pueden no estar identificados a priori.
- Toman mucho tiempo de desarrollo y no tienen la habilidad de actualizarse a medida que exista nueva información de variables
- No realizan cálculos automáticos, se debe ingresar cada variable manualmente, esto implica un alto costo en términos de tiempo

1.2.4. Propuesta de solución

A partir del aumento en el uso de registros electrónicos de salud que se ha llevado a cabo en los últimos años por parte de los establecimientos de salud, sumado al desarrollo técnicas de data mining y machine learning cada vez más sofisticadas, se crea una nueva oportunidad, la de integrar el enorme número de datos disponibles en registros electrónicos, para predecir ciertos desenlaces clínicos en pacientes críticos, desempeñándose como un nuevo tipo de herramienta de apoyo a la decisión, una herramienta basada en datos.

El trabajo propone diseñar y construir un modelo de predicción de mortalidad, en pacientes adultos mayores hospitalizados en Unidades de Cuidados Intensivos del Hospital Clínico de la Universidad de Chile, usando algoritmos de clasificación Data Mining. Su objetivo, es dar solución a la falta indicadores tangibles que revelen la probabilidad de ocurrencia de este, con el fin de entender de mejor manera cuáles son las variables que explican estos resultados, y clasificar a cada paciente según su nivel de riesgo.

1.3. Hipótesis de Investigación

La hipótesis central del trabajo, es que, a partir de los datos creados desde registros electrónicos de salud, correspondientes a información clínica específica de cada paciente, abarcando resultados de laboratorio, fármacos, datos demográficos, diagnósticos y procedimientos, se pueden construir modelos de predecir mortalidad en adultos mayores hospitalizados en unidades de cuidados intensivos.

1.4. Objetivos

1.4.1. Objetivo General

Diseñar y construir un modelo de predicción de mortalidad en pacientes adultos mayores hospitalizados en la Unidad de Cuidados Intensivos del Hospital Clínico de la Universidad de Chile, utilizando algoritmos de Data Mining.

1.4.2. Objetivos Específicos

- Establecer el estado del arte de las técnicas relacionadas a desarrollo y validación de modelos de predicción de desenlaces y pronósticos clínicos en pacientes .
- Construir base de datos correspondiente a pacientes adultos mayores ingresados en la unidad de cuidados intensivos del HCUCH.
- Seleccionar los algoritmos de Data Mining a utilizar que permitan obtener un clasificador de predicción de mortalidad.

- Construir un modelo de clasificación que permita obtener un indicador de riesgo de mortalidad en pacientes críticos adultos mayores.
- Evaluar y concluir en conjunto con contraparte médica en base a los resultados obtenidos.

1.5. Metodología

El primer paso consiste en realizar una revisión bibliográfica de investigaciones que alberguen técnicas relacionadas con el desarrollo y validación de modelos de predicción de desenlaces y pronósticos clínicos.

Luego, se seguirán los pasos de la metodología KDD (*“Knowledge Discovery in Data Bases”*) (Ver figura 2.1).

1. Selección de Datos:

Se selecciona y extrae la data disponible en registros electrónicos de salud del HCUCH, correspondientes a todos los pacientes adultos mayores admitidos en UCI, desde el 01/01/2012 al 31/12/2016. Se identifican tres fuentes de obtención de datos: 1) Ficha Clínica Electrónica 2) Farmacia 3) Laboratorio.

2. Pre-procesamiento y Transformación:

Una vez extraídos, se procede a realizar el pre-procesamiento y la transformación de los datos: la etapa de transformación incluye la transformación de variables continuas en variables discretas en el caso de las variables de laboratorio, y la reducción de la dimensionalidad de la base de datos mediante la categorización de fármacos, diagnósticos y procedimientos. La etapa de pre-procesamiento incluye el tratamiento de valores faltantes en el caso de las variables que no estén presentes para todas las observaciones.

Una vez realizado lo anterior, se cruzan las distintas bases mediante el número de identificación de cada paciente, para consolidar la base de datos a partir de la cual se construye el modelo, la cual incluye las variables predictoras y la variable objetivo: el estado del paciente al alta del hospital, etiquetado en dos clases: una clase positiva, indicador de que el paciente egresa fallecido del hospital, y una clase negativa que indica que el paciente egresa vivo.

3. Data Mining:

Se seleccionan los algoritmos de clasificación y estrategia de entrenamiento para construir y validar el modelo predictivo. Con el objetivo de obtener modelos más simples, rápidos e interpretables, se realiza un análisis univariado y eliminación por correlación para cada una de las variables preliminares, eligiendo un criterio de inclusión según el nivel de significancia de cada atributo con respecto a la variable objetivo, de manera de abarcar solo las variables más relevantes y menos redundantes.

Luego, se entrenan los algoritmos de clasificación seleccionados, que permitirán predecir el desenlace de cada paciente representado por dos clases, con su respectiva probabilidad de ocurrencia, con el objetivo de identificar a cada paciente según un nivel de riesgo.

4. Análisis de Resultados y Conclusiones:

Mediante una validación cruzada repetida de 10 folds, se mide el desempeño de cada clasificador para cada configuración predictiva, según el área bajo la curva ROC (AUC). Se comparan los resultados y se concluye a partir de ellos. Adicionalmente se realiza un análisis sobre la influencia de las variables más relevantes para predecir mortalidad, comparando con lo documentado en la literatura y con el criterio experto mediante la discusión de los resultados con la contraparte del área de la medicina.

1.6. Resultados Esperados

En función de los objetivos del presente trabajo, su desarrollo pretende obtener los siguientes resultados:

- Una base de datos consolidada correspondiente al registro de pacientes ≥ 65 años que ingresaron o estuvieron en la Unidad de Cuidados del HCUCH, durante el período de enero de 2012 a diciembre de 2016.
- Una metodología de selección y aplicación de algoritmos de clasificación de Data Mining en datos electrónicos de salud
- Un modelo de predicción de mortalidad en pacientes adultos mayores hospitalizados en UCI
- Datos estadísticos que permitan probar la relación entre factores clínicos predictores y los desenlaces nombrados en los dos puntos anteriores
- Evaluación y comparativa del modelo nombrado

1.7. Alcances

El foco de este trabajo es predecir mortalidad a partir de datos electrónicos de salud, con el objetivo de caracterizar a cada paciente con un indicador de riesgo y además identificar cuáles son los factores clínicos lo explican. Para esto se utilizarán datos de pacientes adultos mayores chilenos (>65 años a la fecha de admisión), ingresados en UCI del HCUCH entre el 1 de enero de 2012 y el 31 de diciembre de 2016. Los datos corresponden a registros médicos específicos de cada paciente que se extraen de los registros electrónicos del hospital, los cuales corresponden a:

1. Ficha clínica electrónica: La UCI es el único servicio del hospital que no tiene integrada la ficha clínica con el resto de los sistemas de información, por lo cual la extracción de información de esta fuente es limitada. De aquí se desprenden datos demográficos (edad, sexo), las fechas de admisión y alta de hospital y de UCI, tipo de paciente (médico, quirúrgico agendado, quirúrgico urgente), y todos los diagnósticos (principal y secundario) y procedimientos realizados.
2. Farmacia: Corresponde al registro de todos los fármacos y sueros que se solicitan y devuelven al servicio de farmacia por cada paciente, con la respectiva fecha y cantidades.

3. Laboratorio: Corresponde al registro de los exámenes de laboratorio que se solicitan por cada paciente, con su correspondiente resultado y fecha de solicitud.

Con respecto a los datos se precisan las siguientes limitantes:

- Todos los datos fisiológicos del paciente (temperatura, ritmo cardiaco, ritmo respiratorio, presión arterial), son registrados en papel, por lo cual son dejados fuera de este análisis.
- Tanto los diagnósticos y procedimientos corresponden a información de salida del hospital, siendo asignados por el personal de la unidad de ficha clínica y GRD⁶ una vez que el paciente fue dado de alta, a partir de las notas médicas. Esto implica que esta información no está disponible durante el período de hospitalización del paciente.
- Los datos de farmacia corresponden a todos los medicamentos que se solicitaron y devolvieron a la unidad de farmacia durante el período de hospitalización, pero no se registra el consumo efectivo del paciente en la unidad. Esto implica que no se puede saber con certeza ni la cantidad ni el tiempo exactos del consumo.
- Los datos de ingreso y alta del hospital y del servicio están registrados por fecha sin especificar hora, por lo cual es imposible distinguir las horas exactas transcurridas entre el ingreso del paciente a la unidad y cualquier evento clínico asociado a este. Es por esto que para efectos de este estudio se hace la suposición de que cada fecha corresponde a un período de 24 horas.

Además, bajo el contexto de que la generación y validación de este tipo de modelos tiene por objetivo final la construcción de sistemas de apoyo a la decisión, respecto a lo cual queda fuera del alcance de este trabajo:

- El desarrollo de una aplicación funcional basada en el modelo propuesto
- Escalar el modelo en otras instituciones del sistema de salud

1.8. Estructura del Informe

La estructura de este informe es la siguiente:

El presente y primer capítulo trata de la introducción al trabajo de título, abarcando el contexto en el que está inserto y la justificación sobre la relevancia del problema que se plantea. En este contexto se plantean la hipótesis de investigación, los objetivos y la metodología a seguir para llevarlos a cabo.

El segundo capítulo describe el marco teórico y los principales conceptos que se utilizan para llevar a cabo la metodología propuesta por el trabajo.

El tercer capítulo corresponde a una síntesis sobre lo investigado acerca del estado del arte en técnicas y modelos de predicción de mortalidad y estratificación de riesgo en cuidados intensivos.

⁶GRD: Grupos Relacionados con Diagnósticos

El cuarto capítulo abarca el diseño del modelo, partiendo con la descripción del proceso de selección y extracción de datos, para después describir la selección y transformación de estos.

El quinto capítulo describe el proceso de modelamiento de los datos, exponiendo los resultados obtenidos y comparando entre los distintos clasificadores y configuraciones predictivas. Por otro lado, se realiza una evaluación e interpretación de los resultados.

Finalmente, el sexto capítulo corresponde a las conclusiones de la memoria, tratando sobre la metodología y el trabajo realizado, así como de las limitantes y problemas encontrados y el trabajo futuro que se propone a partir de este.

Capítulo 2

Marco Teórico

2.1. Data Mining

2.1.1. KDD y Data Mining

El término “Knowledge Discovery in Databases” fue utilizada por primera vez en 1991 (Piatetsky-Shapiro), con el objetivo de enfatizar que el conocimiento es el producto final de un proceso de descubrimiento basado en datos.

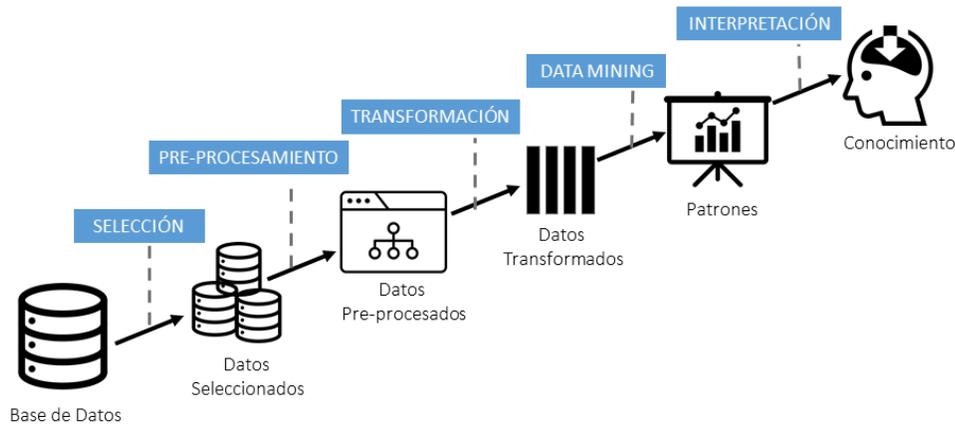
KDD es el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles, y comprensibles en grandes volúmenes de datos [20]. Por otro lado, Data Mining (DM) es un paso esencial del proceso KDD; el cual involucra el desarrollo de algoritmos y modelos matemáticos que exploran la data, con el objetivo de descubrir patrones significativos (implícitos o explícitos), lo cual es la esencia de la gestión del conocimiento.

Es importante entender que Data Mining y KDD son conceptos distintos. KDD se refiere al proceso completo de descubrir conocimiento útil de la data. Data Mining se refiere a descubrir nuevos patrones de un set de datos a través de algoritmos que extraen este conocimiento. (Ver figura 2.1)

Basado en la figura 4, el proceso KDD consiste en la secuencia iterativa de los siguientes métodos:

1. **Selección:** En esta etapa se seleccionan las variables a considerar en el proceso. Para esto se identifican los objetivos del estudio desde el punto de vista del usuario final del conocimiento a descubrir. Después de esto, se selecciona el conjunto de datos y subconjunto de variables y muestras sobre los cuales se procede integrarlos como la base de datos del proceso.
2. **Pre- procesamiento:** Esta es la etapa de análisis y limpieza de los datos. Las operaciones básicas incluyen remover las fuentes de ruido si es apropiado, decidir sobre el uso de las distintas estrategias sobre el tratamiento de valores sin información (missing values),

Figura 2.1: Proceso KDD



Fuente: Elaboración Propia

tratamiento de valores fuera de rango (outliers), etc.

- 3. Transformación:** Esta etapa incluye la reducción y proyección de la data, esto es, encontrar y/o crear atributos útiles que representen y caractericen la data dependiendo del objetivo del proceso. Mediante la reducción dimensional y métodos transformación, el número efectivo de variables pueden ser reducidas, y se pueden descubrir nuevas variantes de representación de la data. Un ejemplo de transformación es el traspaso de una variable continua a nominal, o de una variable nominal a variable binomial, para las cuales se pueden utilizar funciones discretas o continuas.
- 4. Data Mining:** Este paso consiste en el análisis de datos y en la selección de algoritmos que, bajo ciertas limitaciones de eficiencia computacional aceptables, entregan ciertos patrones o modelos sobre la data. Este proceso incluye decidir que modelos y parámetros son los más apropiados según el tipo de datos y variables disponibles, y conciliar un método de Data Mining en particular con el objetivo general del proceso KDD (esto es, por ejemplo, tomar en cuenta si el usuario final está más interesado en entender el modelo por sobre las capacidades predictivas de este). Luego, se aplican los algoritmos para encontrar los patrones modelos o indicadores de interés en una forma representacional en particular.
- 5. Interpretation/Evaluation:** En esta etapa final, se interpretan los patrones encontrados, posiblemente regresando a cualquiera de los pasos anteriores, haciendo de este un proceso iterativo. Además, se seleccionan las medidas de evaluación que permitan decidir las confiabilidad y validez del modelo. Esta etapa puede incluir la visualización de los patrones y modelos, o la visualización de la data dados los modelos extraídos. Luego, viene la evaluación del conocimiento descubierto, ya sea utilizando los nuevos conocimientos directamente, incorporándolo en otros sistemas para trabajo futuro, o simplemente documentando y reportando a las partes interesadas.

2.1.2. Machine Learning

Machine learning es una disciplina del área de las estadísticas y de la inteligencia artificial, cuyo objetivo general es “otorgar a las computadoras la habilidad de aprender sin ser explícitamente programadas”, basándose en la búsqueda de algoritmos que razonan desde instancias externas (datos), para producir hipótesis que luego hacen predicciones sobre instancias futuras [43].

2.1.3. Aprendizaje Supervisado y No Supervisado

El objetivo del aprendizaje supervisado es construir un modelo conciso de distribución de etiquetas de clases, en base a predictores de la base de los datos de entrenamiento. El clasificador resultante, es luego usado para asignar etiquetas de clases a las instancias de prueba (set de prueba), en las cuales los valores del predictor son previamente conocidas, pero el valor de la etiqueta no lo es [43]. Los métodos de aprendizaje supervisado más usados son las técnicas de regresión y las de clasificación.

Métodos de aprendizaje no supervisados son aquellos que no requieren el uso de datos de entrenamiento, es decir, el modelo no es ajustado a las observaciones. Los métodos de aprendizaje no supervisado más usadas son las técnicas de clustering, los cuales buscan agrupar los datos en grupos de objetos similares con el objetivo de reducir la complejidad de la base de datos.

2.2. Missing Data

Un problema común que enfrentan los investigadores al trabajar con técnicas de data mining, es la pérdida de datos cuantitativos. Este es un tema relevante ya que está ampliamente documentado que la aplicación de procedimientos inapropiados de imputación de datos puede introducir sesgos y reducir el poder explicativo de los métodos estadísticos, pudiendo introducir sesgos, pudiendo incluso invalidar los hallazgos y conclusiones.

Según [62], el primer paso a seguir al enfrentar este problema es analizar y definir cuáles son las razones por las cuales existen datos faltantes, y, según esto, sumado al análisis de la distribución de las variables de la base de datos, elegir el método más adecuado para el manejo de los datos faltantes.

2.2.1. Tipos de Valores Faltantes

Según Little y Rubin [52], los mecanismos que llevan a la ausencia de ciertos datos se pueden categorizar en tres grupos:

- **Missing Completely at Random (MCAR):**

Los datos faltantes son MCAR si la razón por la cual cualquiera de las variables contiene ausencia de datos es independiente tanto de la misma variable como de las demás variables del set de datos, ocurriendo de manera completamente aleatoria. MCAR supone que los valores presentes de la variable “X” es una muestra aleatoria de todos los valores de “X”. Entonces, las observaciones con valores perdidos son una muestra aleatoria el conjunto de observaciones, por lo que su distribución es la misma que la de los datos presentes. Generalmente se asumen patrones del tipo MCAR, y es el supuesto que asume la mayoría de los algoritmos de imputación que utilizan los paquetes estadísticos. Sin embargo, en la práctica esto puede ser un supuesto erróneo

- **Missing at Random (MAR):**

La pérdida de datos es MAR cuando los sujetos con datos incompletos son diferentes significativamente de los que presentan datos completos en alguna variable, y el patrón de ausencia de datos puede ser predecible a partir de variables de los registros observados que no muestran ausencia de datos. La probabilidad de que se produzca la ausencia de una observación depende de otras variables del set de datos, pero no de los valores de la variable con el valor ausente.

- **Missing at Random (MAR):**

La pérdida de datos es MNAR cuando la probabilidad de que una variable del set de datos contenga datos perdidos, depende del valor de dicha variable. Esto quiere decir que el hecho de que existan observaciones perdidas contiene información acerca del valor de la variable.

2.2.2. Manejo de Datos Faltantes

La mayoría de los paquetes estadísticos, incluyendo python scikit-learn, no ofrecen algoritmos robustos a set de datos que contengan valores faltantes, razón por la cual, según el patrón de pérdida (descritos en la sección anterior), se debe determinar la forma de manejar la falta de datos. Los métodos utilizados para dar solución a este problema se pueden asociar a tres categorías:

1. **ListWise Deletion**

También conocido como “Complete Case Analysis”, este es el enfoque más simple y probablemente el más utilizado. Asumiendo un patrón de pérdida MCAR, se eliminan todas las instancias que contengan datos faltantes, realizando el análisis únicamente con aquellas que contienen información completa para todas las variables. Esto puede inducir un sesgo en los resultados si los datos no son MCAR, teniendo en cuenta que el tamaño y composición de la muestra fueron elegidas en un procedimiento aleatorio y con probabilidad de selección conocida y distinta de cero, que no puede ser ignorada en el tratamiento de los datos ni en el cálculo de estimadores. Más aún, en el caso de existencia de datos faltantes en registros electrónicos de salud, el primer argumento en contra de este método es que los pacientes con información faltante son sistemáticamente diferentes a aquellos con información completa [74].

2. **Imputación de Datos**

Se han desarrollado varias técnicas de imputación de datos, para los cuales los datos fal-

tantes son rellenados, y la data resultante puede ser analizada como un set de datos completo. Imputar datos en patrones MCAR puede incrementar el poder predictivo, pero no debería cambiar las estimaciones puntuales. En el caso de MNAR, para obtener resultados imparciales (no sesgados), sería necesaria la creación de un modelo que tome en cuenta los datos faltantes que se incorporan en el proceso de imputación [74].

En el caso de registros electrónicos de salud, los valores de otras variables pueden explicar la variación de la ausencia de datos entre pacientes, asumiendo un patrón MAR.

Comúnmente los valores imputados se basan en el valor de los casos completos, por ejemplo, la media o mediana de la variable a imputar o en una estimación del valor de la regresión de otras variables presentes. También se han desarrollado técnicas más sofisticadas tales como la imputación múltiple, rellenando los datos faltantes con más de un valor, con el objetivo de disminuir el sesgo que pueda introducir el proceso de imputación.

Existen también otros enfoques, tales como el estimador de máxima verosimilitud, que, en vez de imputar datos, utilizan procesos iterativos donde se calculan los estimadores del modelo usando sólo los datos completos, y, una vez obtenidos, se estiman los datos faltantes en función del estimador de máxima verosimilitud.

3. Indicador de Ausencia

Todos los métodos descritos en los párrafos anteriores se basan en el supuesto de que los datos perdidos siguen un patrón MCAR o MAR, y no proveen información intrínseca sobre la pérdida de la variable. Según [51], en el caso de registros electrónicos de salud, los mecanismos que llevan a la pérdida o ausencia de los datos poseen información útil para el modelo, es decir, representan datos MNAR, los cuales también son llamados “pérdida no ignorable” o “pérdida informativa”. El mismo autor propone que el mecanismo de ausencia de datos contiene información independiente de los valores observables, se requiere un enfoque de forma de modelar explícitamente la ausencia de la data mediante, por ejemplo, un indicador de ausencia, construido desde una variable dicotómica que representa si la variable fue observada o no.

2.3. Selección de Atributos

Está documentado que en el modelamiento de datos en salud y en bioinformática suele presentarse el problema de alta dimensionalidad, donde muchas veces se tiene un número de atributos mayor que la cantidad de observaciones, también conocido como “curse-of-dimensionality” [8], o small-n-large-p [23]. En consecuencia, si no se realiza la tarea de pre-seleccionar los atributos más relevantes, y no se descartan aquellos que son redundantes, los modelos de predicción tienen un alto riesgo de sobreajustarse [59]. En palabras simples, el sobreajuste u “overfitting”, implica que el proceso de entrenamiento da resultado a un modelo con bajo nivel de generalización, lo que se interpreta como la incapacidad de realizar predicciones precisas sobre nuevas observaciones. Por otro lado, [64] propone que los objetivos más importantes de la reducción de atributos son:

- Evitar sobreajuste y mejorar el desempeño del modelo
- Proveer modelos computacionalmente más rápidos y más costo-efectivos

- Obtener información más profunda acerca de los procesos subyacentes que generaron los datos

En el contexto de clasificación, las técnicas de selección de atributos se pueden agrupar en tres categorías, dependiendo de cómo se combina la selección de atributos con la construcción del modelo:

1. Métodos de Filtro

Estos evalúan la relevancia de los atributos observando solo las propiedades intrínsecas de la data. Lo más común es calcular un puntaje de relevancia para cada atributo en relación a la variable dependiente, removiendo los atributos menos relevantes según un criterio de exclusión. Finalmente, el subset de atributos que queda representa el conjunto de entrada del (de los) algoritmo(s) de clasificación a utilizar.

La ventaja de estos métodos es que son fácilmente escalables a bases de datos de alta dimensionalidad, ya que son computacionalmente simples y rápido, y son independientes del algoritmo de clasificación, de manera que la selección de atributos debe realizarse una sola vez, y después se pueden evaluar distintos algoritmos de clasificación.

Una de las desventajas de estos métodos es que ignoran la interacción con el clasificador. Otra desventaja es que la mayoría corresponde a análisis univariados, lo cual implica que cada variable es considerada de manera separada, ignorando posibles dependencias entre variables. Para mitigar este problema estos métodos combinar con un análisis de correlación.

2. Métodos Wrapper

Estos métodos multivariados utilizan la función objetivo de un modelo de clasificación para jerarquizar los atributos de acuerdo a su relevancia o importancia para el modelo. Estos a su vez se dividen dos sub-categorías: selección hacia adelante (forward selection) y eliminación hacia atrás (backward elimination). Para la primera, la búsqueda de atributos relevantes inicia con un set de variables vacío e iterativamente va agregando variables hasta que se encuentra el subset óptimo de atributos. Por el lado contrario, la búsqueda de eliminación hacia atrás comienza con la inclusión de todos los atributos, que son eliminados iterativamente hasta que se encuentra el número óptimo de atributos.

Las principales desventajas de estos de estos métodos radican en que, por un lado, implican mayor riesgo de sobreajuste que los métodos de filtro, y por otro lado también son mucho más costosos computacionalmente.

3. Métodos Embedded

La búsqueda del subset óptimo de atributos se lleva a cabo en la construcción del clasificador. Al igual que los métodos Wrapper, estos métodos son específicos a un algoritmo dado, teniendo la ventaja de que incluyen la interacción con el modelo de clasificación, siendo a la vez menos costosos.

2.3.1. F-Score

[13] define el puntaje F (o F-score), como una técnica de filtro simple que mide la discriminación de dos sets de números reales. Dados los vectores de entrenamiento x_k , $k = 1, \dots, m$,

si el número de instancias positivas y negativas son n_+ y n_- respectivamente, el puntaje F del i -ésimo atributo se define como:

$$F(i) = \frac{(\tilde{x}_i^{(+)} - \tilde{x}_i)^2 + (\tilde{x}_i^{(-)} - \tilde{x}_i)^2}{\frac{1}{n_+ - 1} \sum_{k=1}^{n_+} (x_{k,i}^{(+)} + \tilde{x}_i^{(+)}) + \frac{1}{n_- - 1} \sum_{k=1}^{n_-} (x_{k,i}^{(-)} + \tilde{x}_i^{(-)})} \quad (2.3.1)$$

Donde \tilde{x}_i , $\tilde{x}_i^{(+)}$ y $\tilde{x}_i^{(-)}$ son el promedio del i -ésimo atributo del set de datos completo, positivo y negativo respectivamente, $x_{k,i}^{(+)}$ es el i -ésimo atributo de la k -ésima instancia positiva, y $x_{k,i}^{(-)}$ es el i -ésimo atributo de la k -ésima instancia negativa. El numerador indica la discriminación entre los sets positivo y negativo, y el denominador indica la discriminación dentro de cada uno de los dos conjuntos.

De esta forma, puntajes F más altos identifican atributos más discriminativos, siendo útil como criterio de selección de atributos. Una vez que se realiza el cálculo, se elige un umbral de significancia (p -valor), que representa la probabilidad de obtener un estadístico mayor que F bajo la hipótesis nula. Así, se selecciona un umbral de significancia óptimo (por ejemplo, $p < 0.05$, $p < 0.1$, $p < 0.2$), que representa una selección de atributos relevantes mediante un proceso de validación cruzada. De esta forma el set de atributos relevantes se utiliza en el modelo de clasificación subsecuente.

La principal desventaja de utilizar este método de selección, es que no toma en cuenta la información mutua entre los atributos, sin embargo, por otro lado, tiene la ventaja de que es simple y generalmente bastante efectivo [13], además es computacionalmente rápido y fácil de implementar, siendo especialmente útil en data de alta dimensionalidad.

2.3.2. Correlación de Pearson

El coeficiente de correlación de Pearson calcula la correlación lineal que existe entre dos atributos, siendo útil para eliminar variables que están altamente correlacionadas, y que, por lo tanto, pueden introducir problemas de redundancia en el modelo. El coeficiente de correlación de Pearson entre dos variables de un modelo se calcula como [59]:

$$P(i) = \frac{\sum_i (x_i - \tilde{x})(y_i - \tilde{y})}{\sqrt{\sum_i (x_i - \tilde{x})^2 \sum_i (y_i - \tilde{y})^2}} \quad (2.3.2)$$

Donde x_i e y_i corresponden al valor de cada atributo para la i -ésima observación, y \tilde{x} y \tilde{y} corresponden al promedio de cada atributo atributos.

Los valores que puede tomar $P(i)$ fluctúan entre -1 y 1, y el nivel de correlación entre dos variables se evalúa considerando el valor absoluto de este, donde valores más altos corresponden a correlaciones más altas.

2.4. Métodos de Clasificación

En data mining, un clasificador es una función que asigna una instancia no etiquetada, a una etiqueta o clase, utilizando las estructuras de datos internas de la instancia. Por otro lado, un inductor, o algoritmo de inducción, construye un clasificador de un conjunto de datos [42].

Existen diversos métodos de clasificación (algoritmos de inducción) basados en distintos campos de investigación, estos se pueden compararse entre sí en base a distintos parámetros [7]:

- Tratamiento de distintos tipos de atributos (categóricos, binarios, continuos, etc)
- Representación del modelo
- Cantidad y reducción de atributos
- Costo computacional
- Interpretabilidad/habilidad de explicar
- Generalización, es decir, se ajustan bien a los casos menos comunes.

A continuación, se enumeran los métodos más comúnmente usados en modelos de predicción.

2.4.1. Regresión Logística

La regresión logística modela la probabilidad logarítmica ("logit"), de una variable binaria Y , utilizando una combinación lineal de covariables (variables independientes) X :

$$\log\left(\frac{P(Y = 1|X)}{1 - P(Y = 1|X)}\right) = X\beta \quad (2.4.1)$$

Esta ecuación puede ser fácilmente reordenada para obtener la probabilidad de ocurrencia de Y :

$$P(Y = 1|X) = \frac{1}{1 + \exp(-X\beta)} \quad (2.4.2)$$

Los coeficientes o pesos de las variable, β son generalmente ajustados utilizando máxima verosimilitud.

Las ventajas de utilizar este clasificador radican en que por un lado es un modelo transparente, que es fácil de interpretar y también familiar para profesionales del área de la medicina, y por otro lado es un modelo que generalmente ofrece buen ajuste y desempeño, comparable con algoritmos más sofisticados.

2.4.2. Naïve Bayes

Naïve Bayes es una técnica de clasificación basada en el Teorema de Bayes, asumiendo independencia entre los atributos predictores. En términos simples, Naive Bayes asume que la presencia de una variable en particular en una clase, no está relacionada con la presencia de cualquiera otra de las variables. A pesar de que estas variables dependan unas de otras, todas estas propiedades contribuyen independientemente en la probabilidad de ocurrencia de cada clase.

Este método tiene la ventaja de que es un modelo simple de construir, de bajo costo computacional, especialmente útil para bases de datos muy grandes. Además, es un modelo fácil de interpretar, y en general otorga un buen ajuste a los datos, incluso comparable con los métodos de machine learning más sofisticados. En términos simples, el Teorema de Bayes propone que la probabilidad posterior de ocurrencia de un evento dado un atributo $P(c|x)$ (en este caso la clase o desenlace que se quiere predecir), es proporcional a la probabilidad previa $P(c)$, por la probabilidad condicional del atributo dado el evento $P(x|c)$, como muestra la ecuación a continuación:

$$P(x|c) = \frac{P(c|x)P(c)}{P(x)} \quad (2.4.3)$$

Donde:

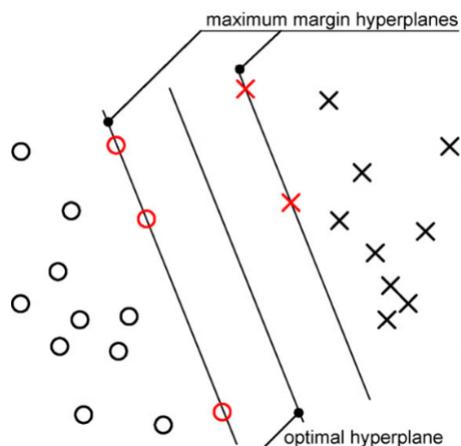
- $P(c|x)$ es la probabilidad posterior de la clase, dado el predictor x
- $P(c)$ es la probabilidad a priori de la clase
- $P(x|c)$ es la verosimilitud, que es la probabilidad del atributo dada la clase
- $P(x)$ es la probabilidad a priori del atributo

2.4.3. Support Vector Machines(SVM)

Las máquinas de soporte vectorial o máquinas de vectores de soporte (Support Vector Machines, SVMs) son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik y su equipo en los laboratorios AT&T [9]. Intuitivamente, un SVM es un modelo que representa a los puntos de muestra en el espacio, separando las clases por un espacio lo más amplio posible. Cuando las nuevas muestras se ponen en correspondencia con dicho modelo, en función de su proximidad pueden ser clasificadas a una u otra clase.

Más formalmente, el algoritmo consiste en la construcción de un hiperplano en un espacio de dimensionalidad muy alta, que separa las muestras de las distintas clases. La idea es encontrar el hiperplano de máxima distancia al punto más cercano de cada clase (hiperplano óptimo); una buena separación entre clases permitirá una clasificación correcta. El set de instancias más cercano al hiperplano óptimo se le llama support vector o vector de soporte 2.2. Una de sus desventajas es que es difícil de interpretar ya que actúa como “caja negra”.

Figura 2.2: Scatterplot de data set de dos clases, con hiperplanos de máxima distancia encontrados mediante algoritmo SVM con núcleo lineal. Las instancias que coinciden en los hiperplanos de máxima distancia (puntos rojos) se llaman vectores de soporte



Fuente: Bellazzi R, Zupan B (2006) *Predictive data mining in clinical medicine: Current issues and guidelines*

Para poder llevar las observaciones a espacios de dimensión mayor, las SVM hacen uso de funciones de enlace o transformaciones, las que se conocen con el nombre de Kernels o núcleo de la función. Entre los tipos de Kernels más utilizados destacan:

- Lineal: $K(x_i, x_j) = x_i^T x_j$
- Polinomial: $K(x_i, x_j) = (\gamma x_i^T x_j)^d, \gamma > 0$
- Función de base radial (RBF): $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$
- Sigmoidea: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

Donde γ , r y d son parámetros exclusivos de cada Kernel.

2.4.4. Redes Neuronales Artificiales

Las Redes Neuronales o *Artificial Neural Networks* (ANN), es una clase de herramienta para el modelamiento de datos que ha gozado de gran popularidad entre investigadores durante los últimos años. Proveniente del campo de la inteligencia artificial, siendo su idea básica la de intentar imitar el aprendizaje de las redes neuronales biológicas, especialmente las del cerebro humano [77].

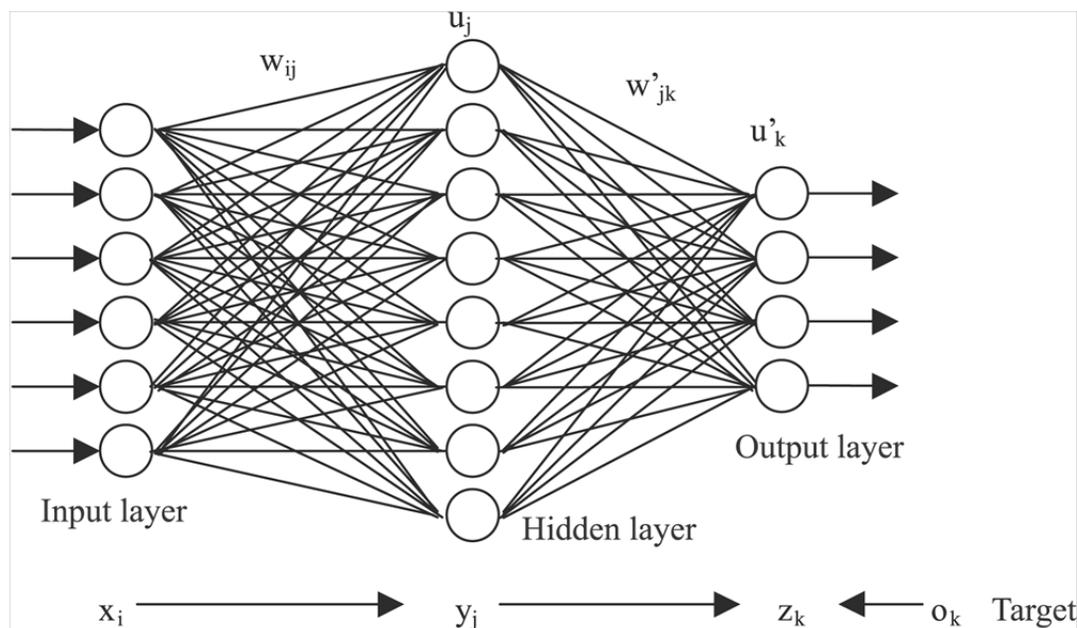
Existen tres tipos de redes neuronales: Multilayer Perceptron (MLP) network, Hopfield network, y Kohonen's map. Para efectos de este trabajo se describe solo la primera, ya que es la que se utiliza para el modelamiento.

MLP es un algoritmo de clasificación que aprende una función $f(\cdot) : R^m \rightarrow R^o$, entrenando sobre una base datos, donde m es el número dimensiones de entrada (número de atributos) y

o es el número de dimensiones de salida. Dado un set de atributos $X = x_1, x_2, \dots, x_m$ y la variable objetivo y , aprende un aproximador de función no lineal. A diferencia de la regresión logística, entre las capas de entrada y salida, pueden existir más capas de neuronas no lineares, llamadas "hidden layers" (capas escondidas). La figura 2.3 muestra la estructura de capas neuronales MLP.

La capa de la izquierda, conocida como la capa de entrada, consiste en un set de neuronas $x_i | x_1, x_2, \dots, x_m$, que representan los atributos de entrada. Cada neurona de la capa oculta transforma los valores de la capa anterior con una suma lineal ponderada $w_1 x_1 + w_2 x_2 + \dots + w_m x_m$, seguido de una función no lineal de activación $g(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$, como por ejemplo sigmoide, hiperbólica, seno o identidad. La capa de salida recibe los valores de la última capa y los transforma en variables de salida.

Figura 2.3: Estructura MLP



Fuente: <https://www.extremetech.com/extreme/215170-artificial-neural-networks-are-changing-the-world-what-are-they>

2.5. Validación de Modelos

Para evaluar el desempeño de un determinado modelo de minería de datos se deben considerar varios aspectos a evaluar:

- Estrategia para el entrenamiento y testeo del modelo
- Tipos de error y medidas de desempeño
- Medidas de ajuste y aplicabilidad

Antes de detallar cada uno de ellos se revisan algunas definiciones básicas [26]:

Definición 2.1 (Conjunto de Datos): *Se distinguen dos tipos, el conjunto de entrenamiento y el conjunto de testeo. Para obtenerlos, se dividen los datos de la muestra de forma aleatoria en dos partes: una de estas se utiliza como conjunto de entrenamiento para construir el modelo y definir los parámetros del clasificador, mientras que la otra parte se utiliza para validar el modelo entrenado y estimar el error de generalización, teniendo en cuenta que el objetivo del clasificador es conseguir un error pequeño evitando el sobreajuste (o “overfitting”), que consiste en la sobrevaloración de la capacidad predictiva del modelo.*

Definición 2.2 (Modelo): *En el caso del presente trabajo también llamado clasificador, es la relación existente entre los atributos predictivos (variables independientes) y el desenlace que se quiere predecir (variable dependiente).*

Definición 2.3 (Error del Modelo): *En los modelos de clasificación, el error del modelo corresponde a la relación de observaciones mal clasificadas sobre el total de observaciones del conjunto de validación. Para estimar el error del modelo existen varias estrategias, detalladas más adelante.*

2.5.1. Entrenamiento y Prueba del Modelo

La idea de partición o “splitting”, es dividir la base de datos en dos conjuntos mutuamente excluyentes, el primero asociado al conjunto de entrenamiento del modelo, y el segundo asociado a la validación de este. Comúnmente se realiza de forma aleatoria a partir de un criterio específico, por ejemplo, usar dos tercios de los datos para el entrenamiento y un tercio para el testeo.

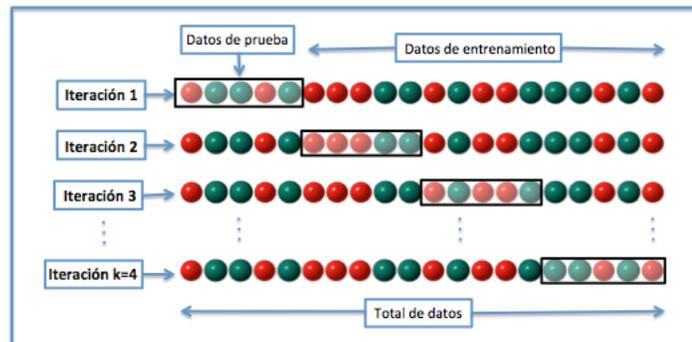
El problema de esto que ambos conjuntos deben ser representativos, es decir, todas las clases deben estar bien representadas en cada conjunto. Al realizar la partición de forma aleatoria existe un trade-off en la proporción de datos a utilizar en cada uno; ya que, por un lado, se debe contar con un número suficientemente grande de observaciones para calibrar el modelo, y por otro lado también es necesario contar con un conjunto de prueba suficientemente robusto y representativo de la muestra, de manera de obtener una estimación de error confiable.

A este método para dividir el conjunto de datos se le conoce como “holdout”. Los parámetros del modelo se obtienen usando el set de entrenamiento, y luego se evalúa el nivel de error utilizando el set de prueba. Pese a su fácil implementación y rapidez computacional, este método implica varias desventajas: al ser un proceso aleatorio, es difícil lograr que ambas particiones sean efectivamente representativas de la muestra inicial, lo que provoca que los resultados obtenidos pueden variar significativamente dependiendo de cómo se seleccionaron las particiones, reflejado en la gran varianza al momento de realizar la evaluación.

Una solución a esto es el método de validación cruzada de n folds (k-fold cross-validation), el cual consiste en la división de la muestra en n subconjuntos disjuntos; el modelo se entrena con $n-1$ conjuntos y se evalúa con el conjunto que se dejó fuera del entrenamiento. Este proceso se realiza n veces (generando n modelos), dejando en cada instancia uno de los subconjuntos como conjunto de validación, como muestra la figura 2.4. De esta manera, el valor de desempeño se calcula como la media aritmética de los valores obtenidos en cada uno de los n modelos.

A excepción de la variante leave-one-out (descrita más adelante en esta sección), la validación cruzada de n folds utiliza una sola partición. Repetir la validación cruzada varias veces usando diferentes particiones en la creación de los folds (o validación cruzada aleatoria) provee una mejor estimación del error, pero implica un costo computacional adicional [42].

Figura 2.4: Esquema de división del conjunto de datos para validación cruzada de 4 folds

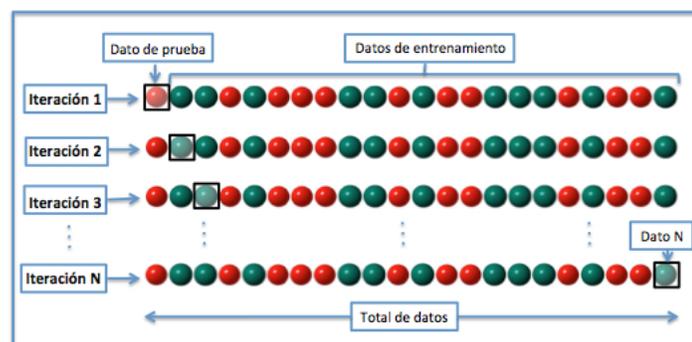


Fuente: https://es.wikipedia.org/wiki/Validación_cruzada

La idea detrás de la validación cruzada, es que el desempeño promedio de los n modelos generados en cada una de la n iteraciones es un buen estimador del desempeño del modelo original (el cual se genera usando todos los datos) al evaluarlo en un futuro conjunto independiente de pacientes [26].

Una variante del método de validación cruzada, es la validación cruzada dejando uno afuera o "Leave-one-out cross-validation" (LOOCV), la cual implica separar los datos de forma que para cada iteración se tenga una sola muestra para los datos de prueba y todo el resto conformando los datos de entrenamiento, como se observa en la figura 2.5. Si bien este método implica mayor costo computacional, la ventaja es que en este tipo de validación cruzada el error es muy bajo, siendo especialmente útil en bases de datos de pocas observaciones.

Figura 2.5: Validación cruzada "Leave one out"(LOOCV)



Fuente: https://es.wikipedia.org/wiki/Validación_cruzada

2.5.2. Tipos de Error y Medidas de Evaluación

Se han desarrollado variadas medidas de error para estimar y comparar el desempeño en modelos de clasificación. Muchas de estas medidas se calculan en función de la matriz de confusión asociada al modelo, la cual se define a continuación:

Definición 2.4 (Matriz de Confusión): La matriz de confusión de un modelo muestra información sobre la clasificación real del set de testeo y la predicha por el modelo. La figura 2.6 representa la matriz de confusión de un clasificador binario.

Figura 2.6: Estructura de una matriz de confusión para problema de clasificación binario

		true class			
		p	n		
predicted class	p'	TP	FP	P'	
	n'	FN	TN	N'	
		P	N		

Fuente: Matjaz Majnik and Zoran Bosnic. *ROC analysis of classifiers in machine learning: A survey*, January 2013

La matriz de confusión contiene la siguiente información:

1. **Verdaderos Positivos:** Número de casos en que el clasificador predijo etiqueta positiva cuando efectivamente poseen etiqueta positiva.
2. **Falsos Positivos:** Número de casos en que el clasificador prefijo etiqueta positiva cuando posee etiqueta negativa.
3. **Verdaderos Negativos:** Número de casos en que el clasificador predijo etiqueta negativa cuando efectivamente poseen etiqueta negativa.
4. **Falsos Negativos:** Número de casos en que el clasificador predijo etiqueta negativa cuando posee etiqueta positiva.

A partir de esta información se construyen las medidas de error comúnmente más ampliamente utilizadas:

Definición 2.5 (Especificidad): Corresponde a la probabilidad de obtener un resultado negativo cuando la instancia está realmente etiquetada como negativa (verdaderos negativos).

En el caso del presente estudio, la especificidad corresponde a la probabilidad de clasificar a un paciente como “vivo al alta” cuando realmente egresó vivo del hospital. Mide la capacidad del modelo de descartar mortalidad o riesgo alto cuando no se manifiesta.

$$\text{Especificidad} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (2.5.1)$$

Definición 2.6 (Sensibilidad): *Corresponde a la probabilidad de obtener un resultado positivo cuando la instancia está etiquetada como positiva (verdaderos positivos).*

En el caso del presente estudio, la sensibilidad corresponde a la probabilidad de clasificar a un paciente como “muerto al alta” cuando el paciente realmente egresó fallecido del hospital. Mide la capacidad del modelo de predecir mortalidad o riesgo alto cuando esta se manifiesta.

$$\text{Sensibilidad} = \frac{\text{TP}}{\text{FN} + \text{TP}} \quad (2.5.2)$$

La *sensibilidad* de un modelo también suele nombrarse en la literatura como *Recall* o TPR (*true positive rate*), mientras que el término FPR (*false positive rate*).

Además de estas dos medidas de error, se pueden considerar las siguientes:

- $\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}}$
- $\text{Precisión o Valor Predictivo Positivo (PPV)} = \frac{\text{TP}}{\text{TP} + \text{FP}}$
- $\text{Valor Predictivo Negativo (NPV)} = \frac{\text{TN}}{\text{TN} + \text{FN}}$

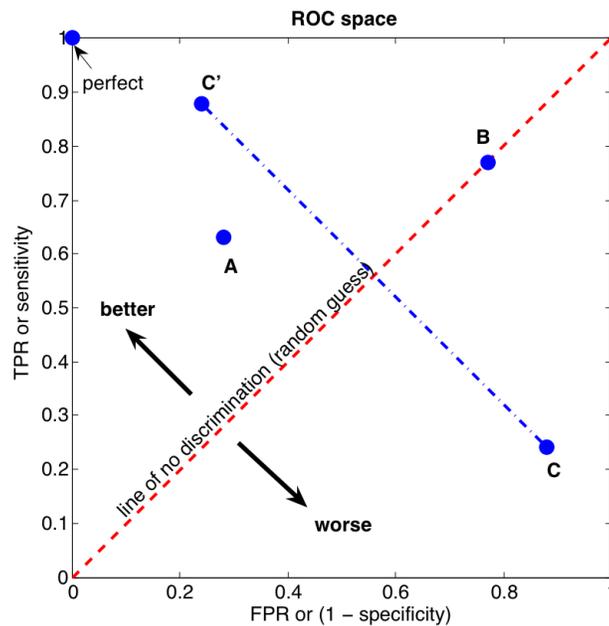
2.5.3. Curva Receiver Operating Characteristics (ROC)

El análisis de curva ROC (Receiver Operating Characteristics), es una metodología para evaluar, comparar y seleccionar clasificadores en base a su desempeño. [54] lo describe como una de las métricas estándar para evaluar algoritmos de machine learning.

En términos concretos, la curva ROC para el problema de clasificación de dos clases, se define como un gráfico de dos dimensiones que representan la tasa de verdaderos positivos (TPR o sensibilidad) en el eje Y, en función de la tasa de falsos positivos (FPR o 1 – especificidad), representada por el eje X. De esta manera, el desempeño de un clasificador, representado por su sensibilidad y especificidad, es representado por un solo punto en el gráfico ROC, como se muestra en la figura 2.7.

Existen algunos puntos característicos básicos en este tipo de gráficos, por ejemplo, el punto de coordenadas (0,0) (TPR=0, FPR=0), representa un clasificador que nunca predice una instancia positiva, mientras que nunca clasifica erróneamente una instancia negativa como positiva. Un clasificador de este tipo no será una buena elección, ya que nunca hará una clasificación correcta de instancias positivas. Por el lado contrario, el punto (1,1) corresponde a un clasificador que etiqueta todas las instancias como positivas, produciendo un gran número de falsos positivos. Ambos puntos nombrados son conocidos como “clasificadores por defecto”. De esta manera, el clasificador perfecto se situaría en el punto (0,1), es decir en la esquina superior izquierda, donde TPR=1 y FPR=0. Si bien no es realista esperar obtener este resultado en problemas de la vida real, representa un punto objetivo a alcanzar en tareas de clasificación.

Figura 2.7: *Espacio ROC*



Fuente: https://commons.wikimedia.org/wiki/File:ROC_space.png

La diagonal ascendente que dividen el espacio ROC representa un clasificador aleatorio. Los puntos por encima de esta diagonal representan buenos resultados de clasificación (mejor que el azar), y los puntos por debajo de la línea corresponde a malos resultados y por lo tanto malos clasificadores, los cuales, sin embargo, podrían ser convertidos fácilmente en buenos predictores invirtiendo las predicciones.

Si bien la curva ROC ha demostrado ser ampliamente útil en la validación de modelos, al momento de comparar dos o más modelos, a medida que la cantidad va aumentando, la comparación visual de las curvas se convierte en una tarea no trivial. Con este propósito se deriva otra medida de ajuste a partir de la curva ROC: el área bajo la curva (AUC), cuyo objetivo es resumir la información que comprende la curva ROC bajo la forma de un valor numérico, facilitando la comparación entre distintos modelos. En un sentido estadístico, el AUC equivale a la probabilidad de que el clasificador evaluará una instancia positiva elegida aleatoriamente, de igual o mejor forma que una instancia negativa aleatoriamente elegida.

Capítulo 3

Modelos de Predicción de Mortalidad

La predicción de mortalidad y la predicción de riesgo han sido ampliamente estudiadas en la literatura. En los últimos años se han desarrollado distintos modelos basados en grandes volúmenes de datos extraídos de registros electrónicos de salud, utilizando algoritmos de machine learning, tales como Regresión Logística, Árboles de Decisión, Support Vector Machines, Redes Neuronales, entre otros [75]. Estos modelos aprovechan la oportunidad dada por la generación y almacenaje de grandes cantidades de información específica de cada paciente hospitalizado, información almacenada bajo la forma de bases de datos que automáticamente capturan data estructurada relacionada a todos los aspectos de la salud, incluyendo diagnósticos, fármacos, resultados de test de laboratorio, imágenes radiológicas, etc. Esta transformación significa una gran promesa para el paciente individual, a medida que información más completa, acoplada a sistemas de apoyo a la decisión, se vuelve fácilmente accesible para apoyar toma de decisiones informadas y así mejorar la seguridad y salud del paciente [39]. Desde la perspectiva de la investigación, poseer data integrada de cada paciente constituye una colección computable de perfiles fenotípicos de grano muy fino, facilitando investigaciones de todo tipo de cohortes, y facilitando el descubrimiento de conocimiento en escalas sin precedentes.

El estado del arte respecto de técnicas predictivas en pacientes críticos se compone de una variada gama de investigaciones, proponen modelos predictivos e indicadores de riesgos basados en diversos desenlaces clínicos: Mortalidad Intra – Hospitalaria, Mortalidad a un período específico (a 30 días, a 6 meses o a 1 año después del alta), Mortalidad en diagnósticos específicos, Readmisión [29].

A continuación se se mencionan diversos modelos de predicción de mortalidad comunes para todo diagnóstico. La tabla 3.1 corresponde a la descripción de los cuatro modelos de clasificación clásicos más ampliamente utilizados en unidades de cuidados intensivos de todo el mundo: APACHE II[41], SAPS II[24], MPM II[50] y SOFA[70].

En la tabla 3.2 se mencionan algunos modelos relevantes de predicción de mortalidad basados en algoritmos de machine learning desarrollados en los últimos años, ordenados por fecha de publicación. Después de hacer una revisión de la literatura respecto de este tipo de modelos, se concluye que hasta el momento se ha llevado a cabo ningún estudio similar a partir de población chilena o latinoamericana.

Tabla 3.1: Sistemas Scoring Clásicos

Modelo (autor/año)	Data Entrenamiento	Predicción (período)	VARIABLES	Método
APACHE II (Knaus 1985)	5815 admisiones UCI de 13 hospitales de EUA	Mortalidad Intrahospitalaria (Primeras 24 horas UCI)	12 variables fisiológicas	Índice escalar entre 0 y 71, basado en sistema asignación de pesos a cada variable en una escala del 0 al 4. Panel de expertos para seleccionar variables y pesos.
SAPS II (Le Gall 1993)	13152 admisiones UCI de 137 hospitales en 12 países de Europa y Norte América (30-sept a 27-dic 1991)	Mortalidad Intrahospitalaria (Primeras 24 horas UCI)	17 variables: 10 fisiológicas, edad, tipo de admisión(3), diagnóstico(3)	Probabilidad de muerte intra-hospitalaria. Para modelamiento se usa regresión logística.
MPM II (Lemeshow 1993)	19124 (MPM0), y 10357 (MPM24) admisiones UCI, de 139 hospitales en 12 países Europa y Norte América (30-sept a 27-dic 1991).	Mortalidad Intrahospitalaria (Admisión UCI (MPM0), y primeras 24 horas (MPM24), pueden ser utilizados por separado o como complemento)	15 variables: 3 fisiológicas, edad, diagnósticos(8), VM, tipo admisión (2)	Probabilidad de muerte intra-hospitalaria. Para modelamiento se usa regresión logística.
(SOFA (Vincent 1996)	No corresponde a un modelo predictivo (no valida en estudio original).	Indicador de fallo multi-orgánico. No predice un desenlace si no que describe una secuencia de complicaciones en pacientes críticos (cada 24 horas).	6 variables: PaO ₂ /FiO ₂ (Respiratorio), Plaquetas (Coagulación), Bilirrubina (Hígado), Hipotensión (Cardiovascular), Glasgow (Sistema Nervioso), Creatinina (Renal).	Puntaje entre 0 y 24 puntos. Se asigna un puntaje de 0 (normal) a 4 (más anormal) para cada sistema, tomando en cuenta los peores valores registrados.

Fuente: Elaboración propia

Tabla 3.2: Modelos predictivos machine learning

Estudio (país)	Data Entrenamiento	Predicción (período)	VARIABLES	Método
Silva 2006 (Portugal) [65]	13164 admisiones UCI adultos de 42 hospitales en 9 países europeos, 11/1998 a 08/1999	Mortalidad UCI	Presencia y duración de eventos adversos + otros (edad, SAPS II, tiempo admisión)	ANN ^a / RL ^b
Hug 2009 (EUA) [36]	10066 admisiones UCI adultos base datos MIMIC II ^c	Mortalidad UCI o a 30 días de alta (Dinámico en tiempo real; 1 a 5 primeros días UCI)	Datos ficha enfermería (signos vitales, laboratorio, tipo UCI), medicamentos, demográficos.	RL
Lehman 2012 (EUA) [49]	14,739 admisiones UCI adultos (MIMIC II)	Mortalidad UCI (Primeras 24 horas)	Notas médicas en texto libre	Topic Modeling + RL
Chen 2014 (EUA) [12]	8000 admisiones UCI adultos (MIMIC II)	Mortalidad UCI (Primeras 48 horas)	Descriptores (edad, sexo, tipo UCI) y ficha clínica (fisiológicos y laboratorio)	Ensemble ANN
Ghassemi 2014 (EUA) [27]	19,308 admisiones adultos UCI (MIMIC II)	Mortalidad UCI (Dinámico en tiempo real 0 a 228 horas)	Notas médicas texto libre + datos estructurados (edad, sexo, SAPS II, dx CIE-9)	Topic modeling + SVM ^d
Nori 2015 (Japón) [60]	1136 admisiones UCI adulto de hospital en Japón, 4/2010 a 3/2013	Mortalidad Hospital (Datos 1 y 2 días previos al alta + modelo retrospectivo)	Dx CIE-10, datos demográficos, eventos clínicos.	Multi-task learning (específicos a cada diagnóstico)
Ball 2016 (Canadá) [60]	1033 admisiones UCI AM (>80a), 22 hospitales en Canadá, 9/2010 a 2/2013	Mortalidad Hospital	VARIABLES APACHE II	RL
Luo 2017 (EUA) [60]	18,412 admisiones UCI adultos (MIMIC II)	Mortalidad a 30 días y 6 meses de alta UCI	Notas médicas texto libre + datos estructurados (edad, sexo, SAPS II, dx CIE-9)	Topic Modeling + SVM

Fuente: Elaboración propia

Capítulo 4

Diseño del Modelo

4.1. Selección y Extracción de Datos

Los datos a utilizar en este trabajo provienen de tres sistemas informáticos utilizados por el HCUCH, independientes entre sí:

4.1.1. Ficha Clínica Electrónica

A diferencia del resto de los servicios del HCUCH, la UCI es el único que aún no integra el sistema de ficha clínica con el resto de los sistemas de información del hospital, por lo cual no cuenta con un sistema que reúna todos los datos de los pacientes (datos fisiológicos, exámenes de laboratorio, diagnósticos, procedimientos, fármacos, etc). Por otro lado, los datos fisiológicos básicos de los pacientes (temperatura, presión arterial, ritmo cardíaco y ritmo respiratorio) son registrados en papel, por lo cual no se encuentran disponibles de forma electrónica, quedando estas variables fuera del modelo propuesto.

Se solicitan todos datos disponibles en ficha clínica electrónica correspondientes a pacientes adultos mayores de 65 años a la fecha de ingreso, que hayan sido admitidos o hayan pasado por la unidad de cuidados intensivos, entre el 1 de enero de 2012 y el 31 de diciembre de 2016. Los datos solicitados son entregados por la Unidad de Ficha Clínica y GRD en un archivo Excel, el cual contiene 424 observaciones correspondientes a cada hospitalización que cumple con los criterios mencionados. En la tabla 4.1 se detallan las variables disponibles en el archivo original además de las variables que se extraen de las originales.

Tabla 4.1: Descripción de atributos set de datos ficha clínica electrónica

Variable	Descripción
ID	Número Identificador del paciente
Edad	Edad del paciente
Sexo	Sexo del paciente
Tipo	Si el paciente es médico, quirúrgico urgente o quirúrgico programado
FIH	Fecha de ingreso hospital
FEH	Fecha de egreso hospital
Servicio de Ingreso	Servicio en que ingresa el paciente
Traslados	Todos los tralados entre servicios del hospital
Servicio de Alta	Servicio del que egresa el paciente
FIU	Fecha de ingreso UCI
FEU	Fecha de egreso UCI
Dias UCI	Días de estadía en UCI
Servicio de Procedencia	Servicio anterior al ingreso a UCI
Diagnóstico Principal	Corresponde a la razón de hospitalización
Diagnósticos Secundarios	Todas las comorbilidades
Procedimientos	Todos los procedimientos aplicados al paciente
Desenlace	Si el paciente egresa vivo o fallecido del hospital

Fuente: Elaboración propia

4.1.2. Farmacia

El sistema de registro electrónico del servicio de farmacia corresponde a una herramienta diseñada y desarrollada exclusivamente para el uso del hospital. A partir del número de ID obtenidos del registro de pacientes entregados por ficha clínica, se extrajeron 415 documentos en formato Excel correspondiente a cada paciente, con el registro de todas las solicitudes o devoluciones de fármacos y sueros desde la UCI al servicio central de farmacia.

Cada solicitud y devolución está identificada por un numero de orden con sus respectivas fechas y cantidades. Cabe mencionar que, ya que el consumo efectivo de fármacos no está registrado en ficha clínica, no existe una completa certeza de que la cantidad exacta de cada fármaco que se solicitó fue efectivamente administrada al paciente.

Para generar la base de datos se extrajo la información de cada uno de los 415 archivos, para juntarlos en un solo archivo en formato csv, registrando el consumo individual para cada paciente, de todos los fármacos que se solicitaron y devolvieron registrados por número de orden y fecha correspondiente, durante el período total de hospitalización. En total se registraron 527 fármacos.

4.1.3. Laboratorio

El sistema de laboratorio es similar al sistema de farmacia, siendo exclusivamente para el uso del HCUCH. En los últimos años en el Servicio Central de Laboratorio se han registrado 244 tipos de exámenes. Por la complejidad y tiempo de extracción de este tipo de datos, para seleccionar que exámenes incluir en el modelo, se realizó una revisión bibliográfica del estado del arte con el objetivo de solicitar solo el set de exámenes que provean factores relevantes para la predecir mortalidad, utilizando como referencia los sistemas clásicos y algunos modelos de predicción de machine learning nombrados en el capítulo 3 [41, 50, 24, 70, 65, 36, 12].

Por otro lado, en esta planificación también se tomó en cuenta la frecuencia de aplicación de cada examen con el objetivo de minimizar los valores faltantes. Finalmente se solicitó un set de exámenes para cada paciente registrado en la base de ficha clínica, el set se detalla en la tabla 4.2. Para cada examen, los datos fueron extraídos en un archivo Excel con el registro de todos los resultados de cada examen realizado en cada paciente durante el período de hospitalización.

Tabla 4.2: Descripción variables exámenes de laboratorio

Variable	Descripción
Electrolitos	Sodio, Potasio, Cloro
Gases y Equilibrio Ácido Base	pH, Bicarbonato Estándar (HCO ₃) y Presión de CO ₂ (Pa-CO ₂)
Hemograma	Hemoglobina, Leucocitos, Plaquetas
Creatinina	Creatinina
Ácido Láctico	Ácido Láctico
Proteína C Reactiva	Proteína C Reactiva
Perfil Bioquímico	Glucosa, Nitrógeno Ureico, Ácido Úrico, Colesterol, Proteínas Totales, Albúmina, Bilirrubina Total, Transaminasa GOT, Deshidrogenasa Láctica (LDH), Fosfatasa Alcalinas, Calcio, Fósforo
Carboxihemoglobina	Carboxihemoglobina, PaO ₂ , Saturación O ₂ (SaO ₂)
Procalcitonina	Procalcitonina

Fuente: Elaboración propia

4.1.4. Base de Datos Preliminar

De las 424 observaciones disponibles en la base de datos de ficha clínica, se aplican tres criterios de exclusión:

- Se eliminan duplicados para aquellos pacientes que fueron hospitalizados dos o más veces en UCI durante el período de estudio. A partir de la opinión médica y de revisión de estudios similares, se decide mantener solo los datos de la primera hospitalización. Se eliminan 9 observaciones.
- Se excluyen todos aquellos que ingresaron a UCI antes del 01/01/2012. Se eliminan 10

observaciones.

- Se excluyen aquellos pacientes de los cuales no se obtuvo el registro de farmacia. Se eliminan 2 observaciones.

Luego de aplicar los criterios de exclusión, la base de datos se compone de 403 observaciones, que corresponden a todos los pacientes de edad igual o mayor a 65 años, que ingresaron o estuvieron en la UCI del HCUCH desde el 1 de enero de 2012 al 31 de diciembre de 2016. Los datos correspondientes a las distintas fuentes de información (Ficha Clínica, Farmacia y Laboratorio), se consolidan mediante un número de identificación de cada paciente.

Para mantener la confidencialidad de la información personal de los pacientes, la información de identificación se anonimiza mediante un número de ID. Cabe mencionar que para la aplicación de técnicas de data mining este campo no tiene relevancia en el análisis que se quiere realizar.

Las variables disponibles en la base de datos se detallan en la tabla 4.3:

Tabla 4.3: Resumen variables extracción inicial

Categoría	Variable	Tipo	Cantidad
Descriptores Generales	Edad	int	1
	Sexo	bin	1
	Tipo UCI	bin	2
	Días pre UCI	int	1
Variables Clínicas	Diagnóstico Principal	bin	220
	Diagnósticos Secundarios	bin	1021
	Procedimientos	bin	508
	Farmacia	bin	527
	Laboratorio	cont	28
Total			2309

Fuente: Elaboración propia

4.1.5. Diseño del cohorte

Además de las variables que se van incluir en el modelo, el diseño de este debe especificar qué período de tiempo en la hospitalización de cada paciente que se va a considerar para predecir su condición al alta. Tanto otros trabajos relacionados como otros sistemas de scoring que se utilizan actualmente en UCI (APACHE II, SAPS II, MPM II) sugieren que las primeras 24 horas de información del paciente son comúnmente las más predictivas para predecir mortalidad. Otros trabajos demuestran que modelos predictivos con configuraciones por debajo de las 72 horas (24, 48 y 72), se desempeñan mejor que para configuraciones de estadía más largas [36]. Por otro lado, usar predicciones retrospectivas (utilizando información del período completo de hospitalización de cada paciente), arroja mejores resultados que utilizar periodos limitados de tiempo [27], pero no son muy útiles para los objetivos que debe cumplir una herramienta

predictiva de apoyo a la decisión, ya que a priori no se sabe qué momento el paciente será dado de alta.

En base a lo anterior, se discute con la contraparte médica, junto a la cual se toma la decisión de que para el desarrollo de este modelo se considerarán tres configuraciones de predicción:

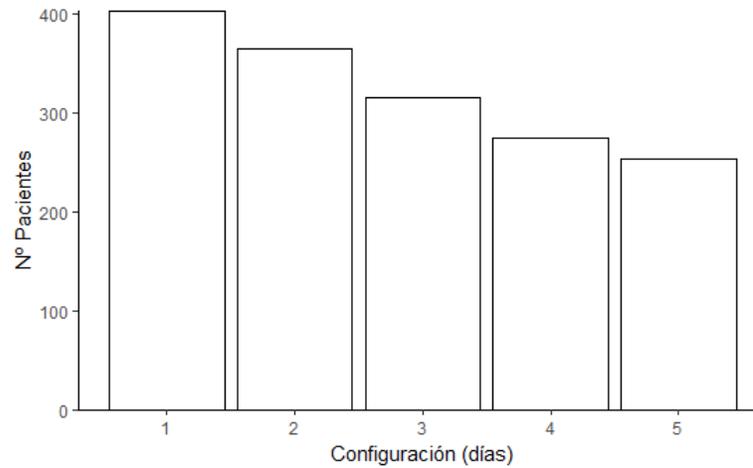
1. Datos primeras 24 horas de hospitalización (403 observaciones)
2. Datos primeras 48 horas de hospitalización (364 observaciones)
3. Datos primeras 72 horas de hospitalización (315 observaciones)

La tarea de predicción difiere según la configuración del set de datos a medida que la data se va acumulando, ya que no todos los pacientes tienen estadías largas (ver figura 4.1), eso significa que mientras se avanza en la línea de tiempo de hospitalización, se cuenta con menos observaciones disponibles y la predicción se vuelve incrementalmente más difícil: para el modelo de 24 horas se cuenta con la totalidad de las observaciones disponibles (403), luego, para el modelo de 48 horas se cuenta con 364 observaciones (disminución del 10% en relación a la base original), y por último, para el modelo de 72 horas se cuenta con 315 observaciones (disminución del 22% en relación a la base original).

Otra ventaja de utilizar configuraciones dinámicas (variables en el tiempo) es que los modelos de este tipo facilitan el desarrollo de sistemas evolutivos de clasificación de riesgo. Esto es especialmente relevante: las UCI son servicios particularmente desafiantes, dado que la gravedad de la enfermedad crítica de cada paciente está constante evolución.

Es importante mencionar que los datos de ingreso y alta del hospital y del servicio están registrados por fecha sin especificar hora, por lo cual es imposible distinguir las horas exactas transcurridas entre el ingreso del paciente a la unidad y cualquier evento clínico asociado a este, específicamente la administración de fármacos y la toma de pruebas de laboratorio. Por esta razón, para efectos de este estudio se hace la suposición de que cada fecha corresponde a un período de 24 horas (o en otras palabras, que el paciente ingresa al servicio las 00:00 de la fecha de ingreso registrada).

Figura 4.1: Cantidad de observaciones disponibles por configuración de predicción



Fuente: Elaboración Propia

4.1.6. Estadística Descriptiva

Con el objetivo de entender mejor la base de datos preliminar, se calculan estadísticas descriptivas para algunas de las variables. Para no extender demasiado, esta sección se limita a entregar un análisis gráfico de un subconjunto de las 2309 variables preliminares descritas en la sección anterior. Figuras 4.2 corresponden a histogramas que describen la distribución de las variables demográficas y descriptores generales, las figuras 4.3 y 4.4 corresponden a la distribución de algunos exámenes de laboratorio.

Figura 4.2: Histogramas variables demográficas y descriptores generales

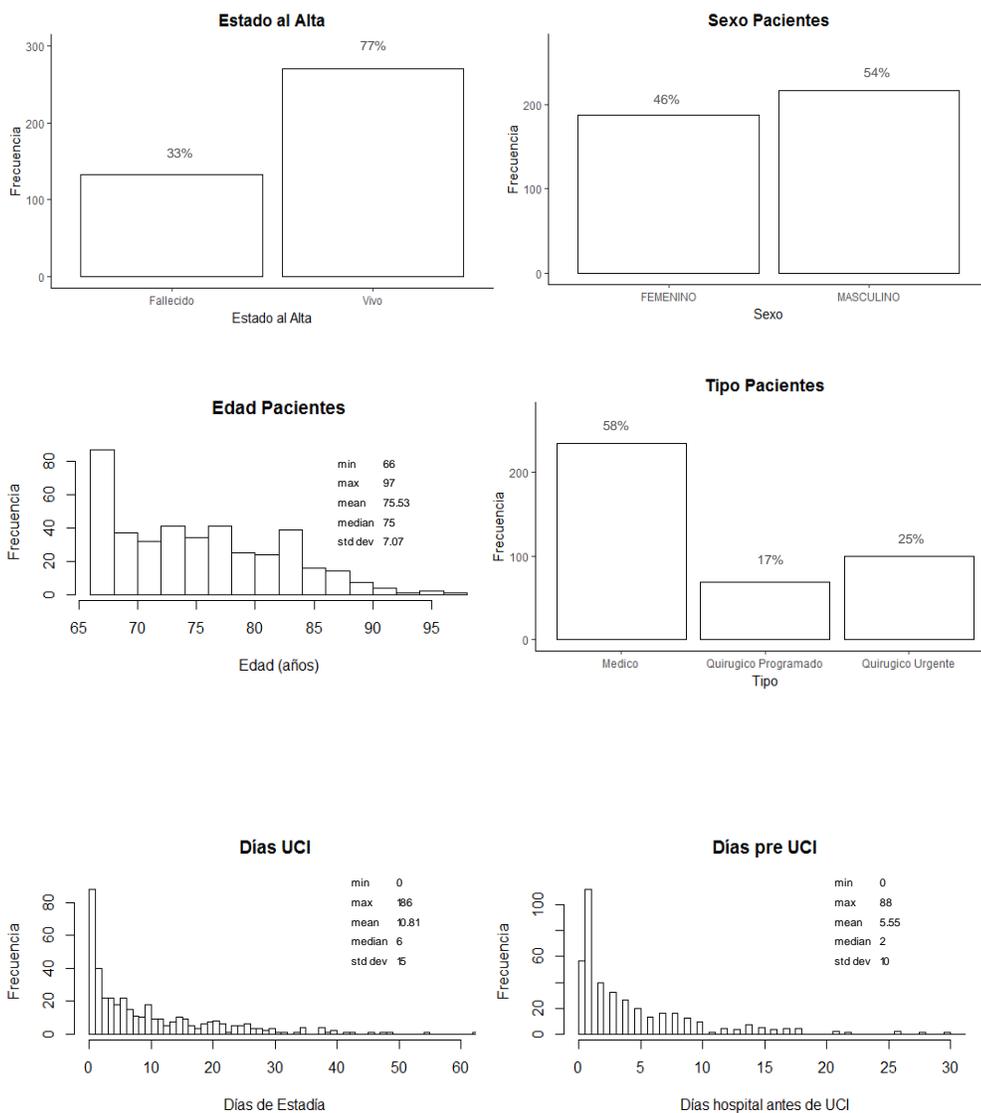
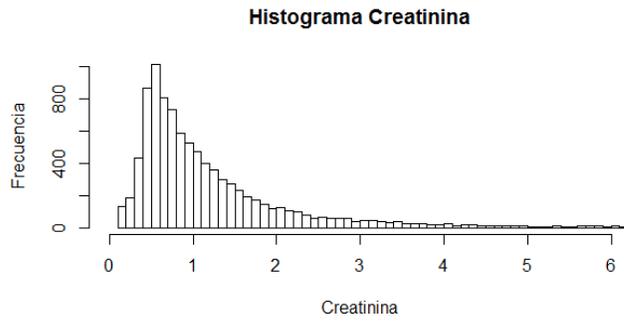
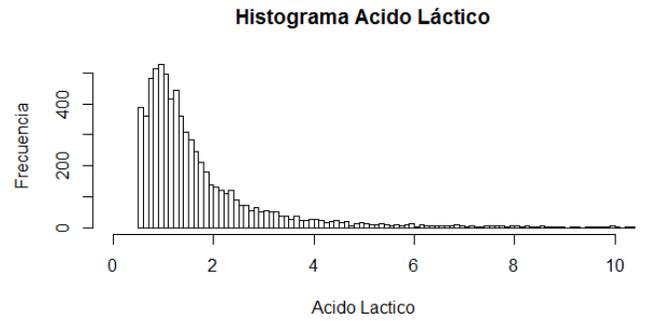


Figura 4.3: Histograma variables laboratorio

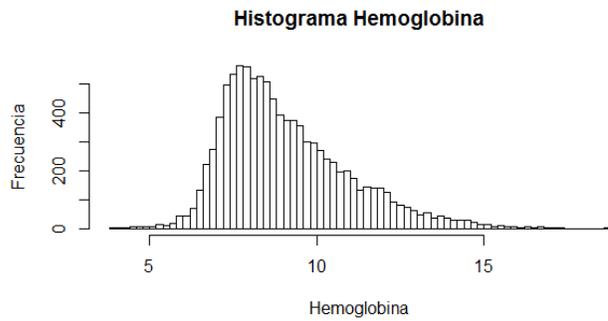
(a) Creatinina



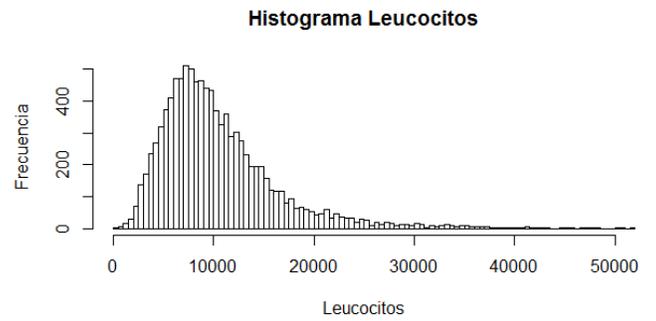
(b) Ácido Láctico



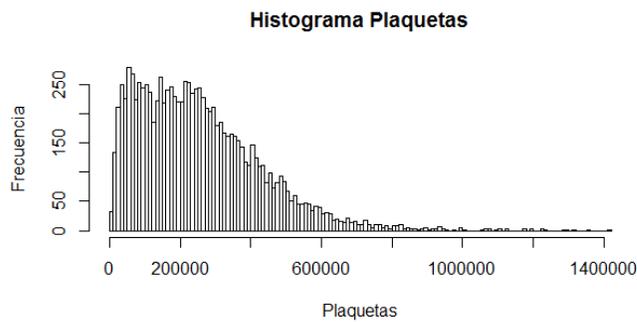
(c) Hemoglobina



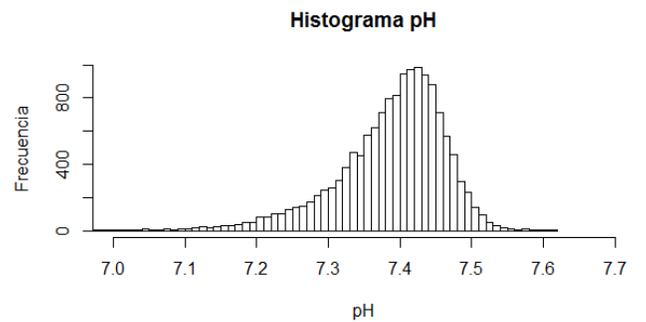
(d) Leucocitos



(e) Plaquetas

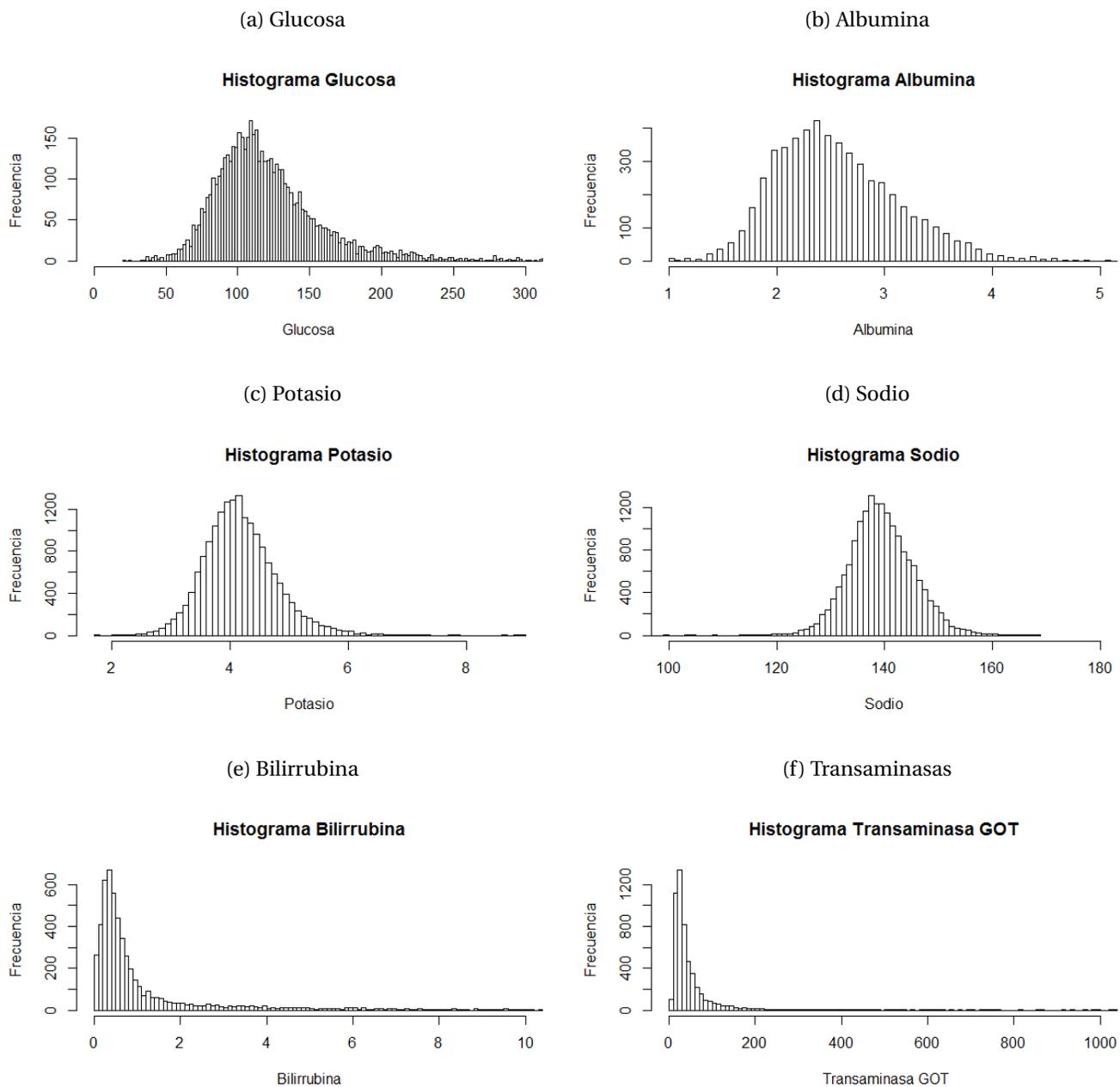


(f) PH



Fuente: Elaboración Propia

Figura 4.4: Histograma variables laboratorio



Fuente: Elaboración Propia

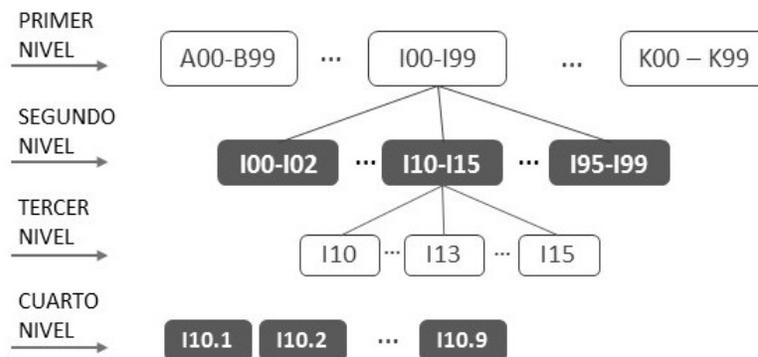
4.2. Selección y Transformación de Variables

4.2.1. Transformación Variables Diagnóstico Principal

El diagnóstico principal corresponde a la razón de hospitalización de cada paciente, por lo cual es información que existe al momento de admisión de cada paciente. Uno de los factores que hace a la UCI unos de los servicios más desafiantes de las instituciones de salud es la diversidad de los pacientes en términos de diagnósticos, teniendo los médicos que enfrentarse a una amplia variedad de enfermedades y patologías. Para la base de datos preliminar se cuenta con 220 distintos diagnósticos.

Los diagnósticos se encuentran disponibles en la base datos codificados bajo el sistema CIE-10. CIE (siglas correspondientes a Clasificación Internacional de Enfermedades en español), es un sistema de clasificación de enfermedades ampliamente utilizado en todo el mundo, el cual fue publicado y es periódicamente actualizado por la Organización Mundial de la Salud. Este corresponde a un sistema jerárquico que categoriza cada diagnóstico en términos de las causas, síntomas, disparidad morfológica, etc, codificando información relevante en cada uno de los niveles de su jerarquía (Ver figura 4.2).

Figura 4.5: Jerarquía CIE 10 representada por cuatro niveles



Fuente: Elaboración Propia

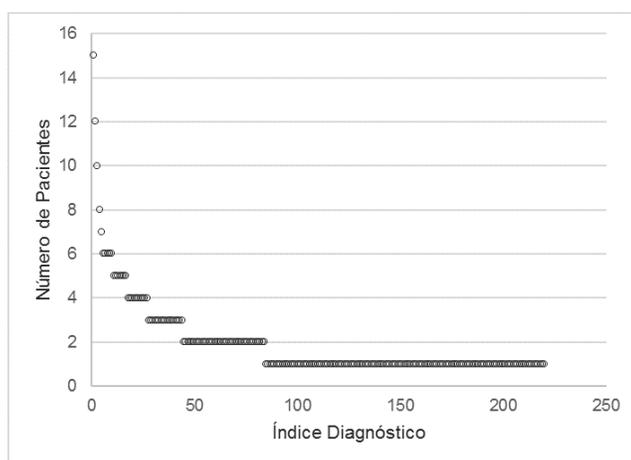
Hoy en día existe un desacuerdo generalizado respecto a la viabilidad de usar los capítulos CIE-10 en tareas predictivas en salud, dado que los códigos diagnósticos fueron creados y publicados para cumplir con propósitos administrativos. Aun así, existe el consenso que al menos en el tercer y cuarto nivel de la jerarquía, la codificación provee una fuente de información y conocimiento generado por expertos de la salud, por lo cual se puede asumir razonablemente que representan un nivel alto de conocimiento experto.

La figura 4.6 muestra el gráfico de distribución de pacientes para cada diagnóstico corres-

pendiente al nivel 4 CIE-10 en la base de datos preliminar: el eje horizontal representa los índices para cada uno de los 220 diagnósticos (representado por un índice numérico), y el eje horizontal corresponde a la cantidad de pacientes. Se puede observar una dispersión importante en los datos, lo que implica que la mayoría de los diagnósticos representan a un número muy reducido de pacientes, y que la mayoría de los diagnósticos cuentan con dos o menos observaciones. Así mismo la figura 4.7 representa la misma distribución utilizando códigos correspondientes al tercer nivel de la jerarquía CIE 10. Se observa que se reduce considerablemente la cantidad de variables (de 220 a 144 diagnósticos), pero sigue existiendo un alto nivel de dispersión en la data y una gran cantidad de diagnósticos que corresponden a dos pacientes o menos.

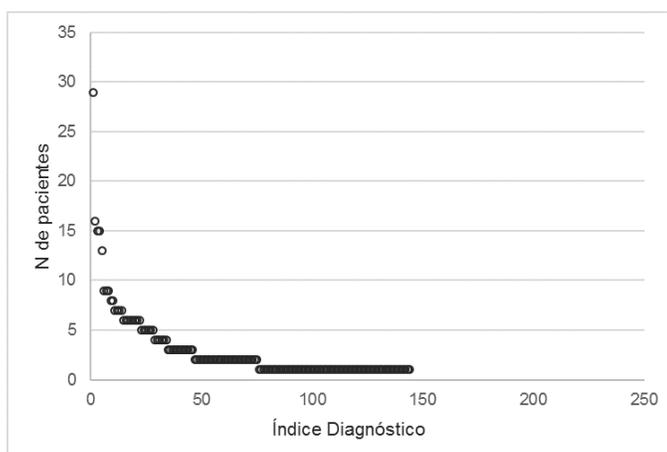
Dado lo anterior, se utiliza la jerarquía CIE para trabajar con el código del tercer nivel, reduciendo así de 220 a 144 variables.

Figura 4.6: Distribución Diagnóstico-Paciente Nivel 4 Jerarquía CIE-10



Fuente: Elaboración Propia

Figura 4.7: Distribución Diagnóstico-Paciente Nivel 3 Jerarquía CIE-10



Fuente: Elaboración Propia

4.2.2. Transformación Variables Diagnósticos Secundarios

Los diagnósticos secundarios de cada paciente corresponden en su mayoría a todas las comorbilidades pre-existentes al momento de admisión, con algunas excepciones, que corresponderían a nuevas patologías que se desarrollan producto del proceso de hospitalización.

Al igual que el diagnóstico principal, los diagnósticos secundarios también están codificados bajo la jerarquía CIE-10. Bajo la hipótesis de que al usar data administrativa en la estratificación de pacientes, en el desarrollo de estudios longitudinales y en el desarrollo de modelos predictivos en salud, las condiciones pre-existentes o comorbilidades deben ser siempre controladas para no provocar ningún tipo de sesgo o pérdida de información del dominio, se han desarrollado algunos sistemas de clasificación que: 1) utilizan indicadores binarios para clasificar la presencia de grupos discretos de comorbilidades para cada paciente, y/o 2) entregan un índice o score que comprende la información de todas las comorbilidades presentes en cada paciente en un solo parámetro.

Dos de los sistemas de clasificación de comorbilidades más ampliamente utilizados en el desarrollo de modelos predictivos son el sistema de Charlson (1987) [11]: el cual propone 17 indicadores binarios que agrupa comorbilidades similares utilizando códigos CIE-9, y un puntaje que compone esta información, y por otro lado, el sistema de Elixhauser (1998) [18], que utiliza 30 categorías de comorbilidades basadas en la codificación del sistema CIE-9, además de un puntaje que comprende la información contenida en las categorías mencionadas.

Quan (2005) [63], y Menendez (2014) [56] muestran evidencia empírica de que el sistema de clasificación de Elixhauser se desempeña mejor que el índice de Charlson en predicción de mortalidad intrahospitalaria. Incluso [63], demuestra que el sistema Elixhauser adaptado a CIE-10 se desempeña igual o mejor que el original CIE-9-CM. Por otro lado, Johnston (2002) [40], plantea que modelos de predicción de mortalidad en UCI, que utilizan grupos de clasificación independientes de Elixhauser, discriminan mejor que modelos que utilizan score APACHE ponderado o el score de Elixhauser, lo cual concuerda lo que concluye Elixhauser [18] en su estudio, respecto de que las comorbilidades tienen efectos independientes en desenlaces clínicos, por lo cual no deberían ser simplificados y reducidos a un solo puntaje, ya que afectan de manera diferente en los resultados entre distintos pacientes.

De acuerdo a lo argumentado en el párrafo anterior, las variables de diagnóstico secundario se transforman utilizando la clasificación de Elixhauser adaptado a CIE-10 para 30 categorías de comorbilidades, utilizando como referencia la traducción de Quan et al [63], que realiza la conversión de la clasificación original (CIE-9-CM) a CIE10. El detalle de cada categoría se detalla en la tabla 4.4 a continuación:

Tabla 4.4: Clasificación de comorbilidades de Elixhauser CIE-10

Categoría	Descripción	Códigos CIE-10
E1	Insuficiencia cardiaca congestiva	I09.9, I11.0, I13.0, I113.2, 125.5, 142.0, 142.5-142.9, I43.x, I50.x, P29.0
E2	Arritmia cardiaca	144.1-I144.3, 145.6, I145.9, I47.x-I49.x, ROO.0, R00.1, R00.8, T82.1, Z45.0, Z95.0

Categoría	Descripción	Códigos CIE-10
E3	Enfermedad valvular cardiaca	A52.0, I05.x-I08.x, I09.1, 109.8, I134.x-I39.x, Q23.0-Q23.3, Z95.2- Z95.4
E4	Enfermedad de la circulación pulmonar	I26.x, I27.x, I28.0, I128.8, 128.9
E5	Enfermedad vascular periférica	I70.x, I71.x, I73.1, I73.8, I73.9, I77.1, I79.0, I79.2, K55.1, K55.8, K55.9, Z95.8, Z95.9
E6	Hipertensión	I10.x I11.x-I13.x, I15.x
E7	Parálisis	G04.1, G11.4, G80.1, G80.2, G81.x, G82.x, G83.0-G83.4, G83.9
E8	Otras enfermedades neurológicas	G10.x-G13.x, G20.x- G22.x, G25.4, G25.5, G31.2, G31.8, G31.9, G32.x, G35.x-G37.x, G40.x, G41.x, G93.1, G93.4, R47.0, R56.x
E9	Enfermedad pulmonar crónica	I27.8, 127.9, J40.x-J47.x, J60.x, J67.x, J68.4, J70.1, J70.3
E10	Diabetes no complicada	E10.0, E10.1, E10.9, E11.0, E11.1, E11.9, E12.0, E12.1, E12.9, E13.0, E13.1, E13.9, E14.0, E14.1, E14.9
E11	Diabetes complicada	E10.2-E10.8, E11.2-E 1.8, E12.2- E12.8, E13.2- E13.8, E14.2-E14.8
E12	Hipotiroidismo	E00.x, E03.x, E89.0
E13	Fallo renal	I12.0, I13.1, N18.x, N19.x, N25.0, Z49.0- Z49.2. Z94.0, Z99.2
E14	Enfermedad hepática	B18.x, I85.x, 186.4, 198.2, K70.x, K71.1, K71.3- K71.5, K71.7, K72.x- K74.x, K76.0, K76.2- K76.9, Z94.4
E15	Úlcera péptica sin sangrado	K25.7, K25.9, K26.7, K26.9, K27.7, K27.9, K28.7, K28.9
E16	Enfermedad por VIH	B20.x-B22.x, B24.x
E17	Linfoma	C81.x-C85.x, C88.x, C96.x, C90.0, C90.2
E18	Cáncer metastásico	C77.x - C80.x
E19	Tumor sólido sin metástasis	C00.x-C26.x, C30.x-C34.x, C37.x-C41.x, C43.x, C45.x-C58.x, C60.x-C76.x, C97.x
E20	Artritis reumatoide	L94.0, L94.1, L94.3, M05.x, M06.x, M08.x, M12.0, M12.3, M30.x, M31.0- M31.3, M32.x- M35.x, M45.x, M46.1, M46.8, M46.9
E21	Coagulopatía	D65-D68.x, D69.1, D69.3- D69.6
E22	Obesidad	E66.x
E23	Pérdida de peso	E40.x-E46.x, R63.4, R64
E24	Alteración de fluidos y electrolitos	E22.2, E86.x, E87.x
E25	Anemia por pérdida de sangre	D50.0
E26	Anemia por otras deficiencias	D50.8, D50.9, D51.x-D53.x
E27	Alcoholismo	F10, E52, G62.1, I142.6, K29.2, K70.0, K70.3, K70.9, T51.x, Z50.2, Z71.4, Z72.1
E28	Abuso de drogas	F11.x-F16.x, F18.x, F19.x, Z71.5, Z72.2

Categoría	Descripción	Códigos CIE-10
E29	Psicosis	F20.x, F22.x-F25.x, F28.x, F29.x, F30.2, F31.2, F31.5
E30	Depresión	F20.4, F31.3-F31.5, F32.x, F33.x, F34.1, F41.2, F43.2

Fuente: Elaboración propia

4.2.3. Transformación Variables Procedimientos

En la base de datos preliminar se dispone de 508 variables correspondientes a procedimientos codificados en sistema CIE-10.

Si bien esta es información valiosa y potencialmente útil en tareas de predicción que utilicen un set de datos retrospectivo o en predicciones de mortalidad de largo plazo, en el caso de este modelo, su incorporación puede producir sesgos ya que en la base de datos preliminar no se especifica en qué marca de tiempo o período de la hospitalización del paciente se realiza cada procedimiento que se le atribuye. Por esta razón del total de variables disponibles la única que se incluye en el modelo es “Ventilación Mecánica”. La razón selección de esta variable es que, a criterio experto y según la literatura [24, 50, 12, 36], es una variable relevante para predecir mortalidad.

4.2.4. Transformación Variables Farmacia

Según [31], existen distintos criterios de clasificar y agrupar fármacos y medicamentos, siendo las más comunes:

1. Fármacos relacionados por su estructura química
2. Fármacos relacionados por su mecanismo de acción
3. Fármacos que se utilizan para el mismo propósito

Por otro lado, según [60], el hecho de que un grupo de fármacos compartan dos o más de estos criterios, incluyendo su eficacia, sugiere que dichos fármacos deben jugar roles similares en la predicción de riesgo del paciente que los consume. Según esta información, se realiza un trabajo de investigación con médicos intensivistas en conjunto con el químico-farmacéutico de la UCI del HCUCH, se crearon seis grupos o clases de fármacos que comparten tanto su propósito, eficacia y mecanismo de acción (ver tabla 4.5). Como se menciona en la sección 4.1.2, como no se tiene certeza de la cantidad exacta administrada efectivamente al paciente, se utilizan indicadores de consumo, de forma que cada categoría es representada una variable binaria que toma valor verdadero (TRUE), si es que el paciente consumió una o más veces, uno o más de los fármacos pertenecientes a dicha categoría para cada período de estudio (24, 48 y 72 horas).

Tabla 4.5: Clasificación de Medicamentos

Grupo	Medicamentos
Profilaxis Gástricos	Esomeprazol, Famotidina, Omeprazol, Lanzoprazol, Ranitidina
Profilaxis Tromboembólicos (Anticoagulantes)	Dalteparina Sodica, Heparina Sodica
Sedantes	Midazolam, Dexmedetomidina Clorhidrato, Fentanila, Propofol, Solución Fenta
Relajantes Neuro-Musculares (RNM)	Atracurio Besilato, Cisatracurio Besilato, Rocuronio, Rocuronio Bromurio, Succinil Colina Cloruro
Vasoactivos	Adrenalina, Dobutamina, Dopamina, Fenilefrina, Milrinona, Norepinefrina
Anticonvulsivantes	Acido Valproico, Fenotoina Sodica, Levetiracetam, Propofol, Amiodarona, Epinefrina, Nitroglicerina, Nitroprusiato, Levosimendan

Fuente: Elaboración propia

4.2.5. Transformación Variables Exámenes de Laboratorio

Como se describe en la sección 4.2.3, la base de datos preliminar comprende 9 exámenes de laboratorio que derivan en 28 variables descritas en la tabla 4.2.

En el set de datos preliminar, para cada una de las variables de laboratorio mencionadas, cada paciente cuenta con un set de observaciones que corresponden a los resultados obtenidos después de cada examen solicitado durante el período de hospitalización. Cada resultado tiene asociado su correspondiente fecha de solicitud. Para efectos del modelo, se debe transformar el set de datos con el objetivo que cada paciente sea representado por uno o más valores que entreguen la mayor cantidad de información respecto al valor y evolución de cada una de las variables de laboratorio.

Por otro lado, después de realizar el análisis exploratorio de estas variables (figuras 4.3 y 4.4), se puede observar que la mayoría, o bien distribuyen normalmente (glucosa, fosforo, cloro, calcio, sodio, potasio, albumina, carboxihemoglobina, HCO, PaO₂, PaCO₂, proteínas totales, SaO₂), o bien según distribución log-normal (ácido úrico, ácido láctico, bilirrubina, creatinina, fosfatasas alcalinas, hemoglobina, leucocitos, plaquetas, procalcitonina, transaminasas, deshidrogenasa láctica, nitrógeno ureico, pH), a excepción de proteína c reactiva (PCR) y colesterol. Dado que no todas las variables distribuyen normalmente, se considera que debe haber especial cuidado en la forma en que se agregar la información, ya que, por ejemplo, caracterizar al paciente solamente según el promedio de resultados puede producir sesgos y falta de información.

Dado lo anterior, y tomando como referencia el modelo de Chen et al [12], se considera representativo transformar la data de manera de atribuir cinco valores a cada paciente, por cada variable: valor de ingreso, valor promedio, valor mínimo, valor máximo y último valor registrado.

Las variables se resumen en la tabla 4.6 a continuación:

Tabla 4.6: Variables derivadas resultados laboratorio

Variable	Descripción
Ingreso	Primera observación registrada durante el período de predicción (24, 48 o 72 horas)
Máximo	Máximo valor entre todas las observaciones registradas durante el período de predicción (24, 48 o 72 horas)
Mínimo	Máximo valor entre todas las observaciones registradas durante el período de predicción (24, 48 o 72 horas)
Promedio	Promedio de todas las observaciones registradas durante el período de predicción (24, 48 o 72 horas)
Último	Última observación registrada durante el período de predicción (24, 48 o 72 horas)

Fuente: Elaboración propia

4.3. Manejo de Datos Faltantes

De todos los atributos predictores descritos anteriormente, los únicos que contienen datos faltantes son los que corresponden a exámenes de laboratorio; el nivel de completitud de cada examen para cada configuración de predicción se muestra en la tabla 4.7. El porcentaje de completitud de cada examen corresponde a la proporción de pacientes que tiene disponible una o más mediciones durante las primeras 24 48 y 72 horas de hospitalización en cada caso, sobre el total de pacientes de la muestra para cada configuración (403, 364 y 315 respectivamente). De la tabla se puede observar que a medida que transcurre el tiempo de hospitalización, aumenta el nivel de completitud para todos los exámenes; esto ocurre porque a medida que transcurre más tiempo de hospitalización cada paciente tiene más probabilidades de tener un mayor número de mediciones disponibles.

Tabla 4.7: Portcentaje de complitud variables de laboratorio por configuración de predicción

Examen	% Completitud 24	% Completitud 48	% Completitud 72
Creatinina	83,37%	93,13%	95,24%
Ácido Láctico	82,63%	90,66%	92,38%
Electrolitos	92,56%	95,33%	96,19%
Gases	91,56%	95,33%	96,19%
Hemograma	83,87%	93,41%	95,24%
PCR	69,48%	84,34%	89,21%
Perfil Bioquímico	70,47%	85,99%	92,70%
Carboxihemoglobina	67,00%	76,65%	80,32%
Procalcitonina	54,09%	70,33%	79,68%

Fuente: Elaboración propia

Ya que no existen paquetes Python que incluyan clasificadores robustos a datos ausentes, se debe decidir que técnica de manejo de datos faltantes utilizar. El método más común para este objetivo es mediante la técnica listwise deletion, la cual elimina todos aquellos casos que tienen algún valor ausente para cualquiera de los atributos. En el caso del presente modelo esto sería incorrecto ya que por un lado se estaría eliminando más de la mitad de las observaciones disponibles para predecir mortalidad a partir de las primeras 24 horas, y que, por otro lado, es correcto aplicar esta técnica solo si los datos ausentes siguen un patrón MCAR, lo cual en este caso no se puede asumir, ya que, según el criterio experto de la contraparte médica y según [74], la razón por la cual no se le pide un examen a un paciente no es aleatoria, sino que depende del estado de salud del paciente. A modo de ejemplo, se puede asumir que es más probable que no se solicite el examen de proteína c reactiva (PCR) a un paciente que no presenta signos de inflamación en el cuerpo.

Bajo esta lógica, sería correcto asumir que los mecanismos que conducen a la ausencia de una medida de laboratorio, contiene información acerca del estado de salud del paciente, es decir, siguen un patrón de pérdida MNAR. Por esta razón, para manejar los datos faltantes, se decide crear una nueva variable binaria para cada variable con datos ausentes, que indica si la variable fue registrada o no. Como los valores de exámenes de laboratorio son originalmente continuos, para que este método sea consistente con el tipo de variable que se va a entregar al modelo, cada una de las variables mencionadas se discretiza en variables nuevas que representan rangos críticos de normalidad (ver tabla 4.8). Los valores de rangos normales utilizados para la construcción de las variables fueron proporcionados por el servicio de laboratorio central del HCUCH, mientras que el resto de los umbrales críticos fueron sugeridos por el equipo médico.

Tabla 4.8: Variables de rango laboratorio

Examen	Cod	Variables	Rango	Unidad
Creatinina	CR	Bajo Rango Normal Sobre Rango Normal 1 Sobre Rango Normal 2	$CR < 0.7/0.8$ (F/M) $1.2/1.5 < CR \leq 2$ (F/M) $CR > 2$	mg/dl
Aclaramiento Creatinina	CRC	Bajo Rango Normal 1 Bajo Rango Normal 2 Bajo Rango Normal 3	$CRC < 30$ $30 \leq CRC < 45$ $45 \leq CRC < 60$	ml/min
Ácido Láctico	AL	Bajo Rango Normal Sobre Rango Normal 1 Sobre Rango Normal 2	$AL < 0,7$ $2,1 < AL \leq 4$ $AL > 4$	mmol/L
Sodio	NA	Bajo Rango Normal Sobre Rango Normal	$NA < 137$ $NA > 145$	mEq/L
Potasio	K	Bajo Rango Normal Sobre Rango Normal	$K < 3,5$ $K > 5,1$	mEq/L
Cloro	NA	Bajo Rango Normal Sobre Rango Normal	$CL < 98$ $CL > 107$	mEq/L
Ph	PH	Bajo Rango Normal 1 Bajo Rango Normal 2 Sobre Rango Normal 1 Sobre Rango Normal 2	$PH < 7,25$ $7,25 \leq PH < 7,37$ $7,42 < PH \leq 7,5$ $PH > 7,5$	

Examen	Cod	Variables	Rango	Unidad
pCO2	PCO2	Bajo Rango Normal Sobre Rango Normal	PCO2<38 PCO2>42	mmHg
HCO3	HCO3	Bajo Rango Normal Sobre Rango Normal	HCO3<23 HCO3>25	mEq/L
Hemoglobina	HGB	Bajo Rango Normal 1 Bajo Rango Normal 2	HGB<7 7≤HGB<12/13.5 (F/M)	gr/dl
Leucocitos	WBC	Bajo Rango Normal 1 Bajo Rango Normal 2 Sobre Rango Normal 1 Sobre Rango Normal 2	WBC<2000 2000≤WBC<4000 10000<WBC≤15000 WBC>15000	K/ul
Plaquetas	PLT	Bajo Rango Normal 1 Bajo Rango Normal 2 Sobre Rango Normal 1 Sobre Rango Normal 2	PLT<80000 80000≤PLT<150000 400000<PLT≤600000 PLT>600000	K/ul
PCR	CRP	Sobre Rango Normal 1 Sobre Rango Normal 2 Sobre Rango Normal 3	100<CRP≤150 150<CRP≤200 CRP>200	mg/L
Glucosa	GLC	Bajo Rango Normal Sobre Rango Normal 1 Sobre Rango Normal 2 Sobre Rango Normal 3	GLC<75 100<GLC≤150 150<GLC≤200 GLC>200	mg/dl
Nitrógeno Uréico	BUN	Bajo Rango Normal Sobre Rango Normal 1 Sobre Rango Normal 2	BUN<7/9 (F/M) 17/20<BUN≤40 (F/M) BUN>40	mg/dl
Ácido Úrico	UA	Bajo Rango Normal Sobre Rango Normal	UA<2,5/3,5 (F/M) UA>7,5/8,5 (F/M)	mg/dl
Calcio	CA	Bajo Rango Normal Sobre Rango Normal	CA<8,4 CA>10,2	mg/dl
Fósforo	P	Sobre Rango Normal 1 Sobre Rango Normal 1 Sobre Rango Normal 2	P<2,5 4,5<P≤5,5 P>5,5	mg/dl
Proteínas	PROT	Bajo Rango Normal Sobre Rango Normal	PROT<6,3 PROT>8,2	g/dl
Albumina	ALB	Bajo Rango Normal 1 Bajo Rango Normal 2 Bajo Rango Normal 3	ALB<2 2≤ALB<2,8 2,8≤ALB<3,5	g/dl
Colesterol	CHO	Bajo Rango Normal Sobre Rango Normal	CHO<50 CHO>200	mg/dl
Bilirrubina	BLR	Bajo Rango Normal Sobre Rango Normal 1 Sobre Rango Normal 2	BLR<0,2 1,3<BLR≤2 BLR>2	mg/dl
Transaminasas GOT	AST	Bajo Rango Normal	AST<14/17 (F/M)	U/L

Examen	Cod	Variables	Rango	Unidad
		Sobre Rango Normal 1	36/59<AST≤72/118 (F/M)	
		Sobre Rango Normal 2	72/118<AST≤126/206 (F/M)	
		Sobre Rango Normal 3	AST>126/206 (F/M)	
Fosfatasas Alcali- nas	ALP	Bajo Rango Normal	ALP<38	U/L
		Sobre Rango Normal 1	126<ALP≤252	
		Sobre Rango Normal 2	ALP>252	
LDH	LDH	Bajo Rango Normal	LDH<313	U/L
		Sobre Rango Normal	LDH>618	
Carboxi hemo- globina	COHG	Bajo Rango Normal	COHG<0,5 %	%
		Sobre Rango Normal	COHG>2,5 %	
pO2	PO2	Bajo Rango Normal 1	PO2<30	mmHg
		Bajo Rango Normal 2	30≤PO2<40	
		Bajo Rango Normal 3	40≤PO2<50	
sO2	SO2	Bajo Rango Normal 1	SO2<60 %	%
		Bajo Rango Normal 2	60 %≤SO2<70 %	
		Bajo Rango Normal 3	70 %≤SO2<80 %	
Procalcitonina	PCT	Sobre Rango Normal 1	0,5<PCT≤2	ng/L
		Sobre Rango Normal 2	2<CRP≤4	
		Sobre Rango Normal 3	CRP>4	

Fuente: Elaboración propia

Capítulo 5

Modelamiento

En el presente capítulo se describen las técnicas de modelamiento utilizadas para predecir la mortalidad intra-hospitalaria para las tres configuraciones de predicción elegidas: primeras 24 horas, 48 horas y 72 horas de hospitalización en UCI.

En primer lugar, en la sección 5.1 se describe la selección de atributos que se lleva a cabo previa al entrenamiento para cada una de las configuraciones.

Luego, en las secciones 5.2, 5.3, y 5.4 se detallan los resultados del modelamiento para cada configuración, 24, 48 y 72 horas respectivamente, mediante el entrenamiento de cuatro algoritmos de clasificación; dos de ellos son modelos definidos como inductores clásicos, simples e interpretables: Regresión Logística (Logit) y Naive Bayes (NB), y por otro lado dos modelos "caja negra": Support Vector Machines (SVM) y Redes Neuronales (ANN), definidos como algoritmos más complejos que a su vez son más difíciles de interpretar.

Además, para cada modelo se reportan los coeficientes dados por el entrenamiento del clasificador Logit para las variables más relevantes, a modo de interpretar cuáles son los procesos y factores asociados a menor o mayor riesgo en pacientes admitidos en UCI. Para el análisis, los coeficientes positivos indican que la variable es un factor asociado a mayor riesgo de fallecer durante la hospitalización, mientras que los coeficientes negativos indican que la variable es "protectora", es decir, se asocia a menores riesgos de que el paciente fallezca y por lo tanto a mayores probabilidades de obtener resultados más favorables.

Finalmente, en la sección 5.5 se realiza una evaluación y comparación de los distintos modelos.

5.1. Selección de Variables

Después de realizar las transformaciones detalladas en el capítulo 4, el set de datos se compone de 735 variables, las cuales se resumen en la tabla 5.1 a continuación:

Tabla 5.1: Resumen variables transformadas

Categoría	Variable	Tipo	Cantidad
Descriptores Generales	Edad	int	1
	Sexo	bin	1
	Tipo UCI	bin	2
	Días pre UCI	int	1
	Ventilación Mecánica	bin	1
Variables Clínicas	Diagnóstico Principal	bin	144
	Diagnósticos Secundarios	bin	30
	Farmacia	bin	6
	Laboratorio	bin	549
Total			735

Fuente: Elaboración propia

Podemos notar que, para todas las configuraciones de predicción, el número de covariables predictoras es considerablemente mayor que el número de observaciones: 735 covariables versus 403, 364 y 315 observaciones para modelos de 24, 48 y 72 horas respectivamente. Esto está ampliamente documentado en la literatura, nombrado también como “curse-of-dimensionality” [8], o small-n-large-p [23], y que puede inducir sobreajuste en los modelos disminuyendo su potencial desempeño, además de altos costos computacionales, y baja calidad en la información acerca de los procesos que generaron los datos.

Es por eso que, con el objetivo de disminuir la dimensionalidad de la base de datos, de manera de considerar en el modelo solo las más discriminativas y menos redundantes, se lleva a cabo un análisis univariado sobre las variables candidatas:

Análisis Univariado:

Dada la variable objetivo binaria que se quiere predecir, se realiza un análisis univariado mediante la aplicación de un test F de significancia, ordenando las variables según su relevancia para el modelo.

Una vez creado el ranking, se elige el umbral de corte para la selección de variables en función del p-valor de cada covariable (ej: $p=0.05$), de manera de hacer la distinción entre puntajes altos y bajos, e incluir en el modelo solo las variables cuyo p-valor sea menor que el umbral elegido. Luego, para cada par de variables se calcula el coeficiente de correlación de Pearson (ρ), eliminando así covariables altamente correlacionadas. En la práctica este proceso lleva a la selección de una sola de las transformaciones realizadas sobre cada medida de laboratorio (promedio, mínimo, máximo, ingreso o último registro). Además, para cada una de las variables continuas (edad y días pre UCI), se realiza una transformación logarítmica para luego evaluar cuál es más relevante según el puntaje F.

La elección del umbral p-valor utilizado para la selección de variables no es un proceso trivial: muchos autores sugieren el uso de umbrales más holgados (ej: $p=0.25$) con el argumento de que así el modelo puede capturar el efecto conjunto o información mutua de las covariables,

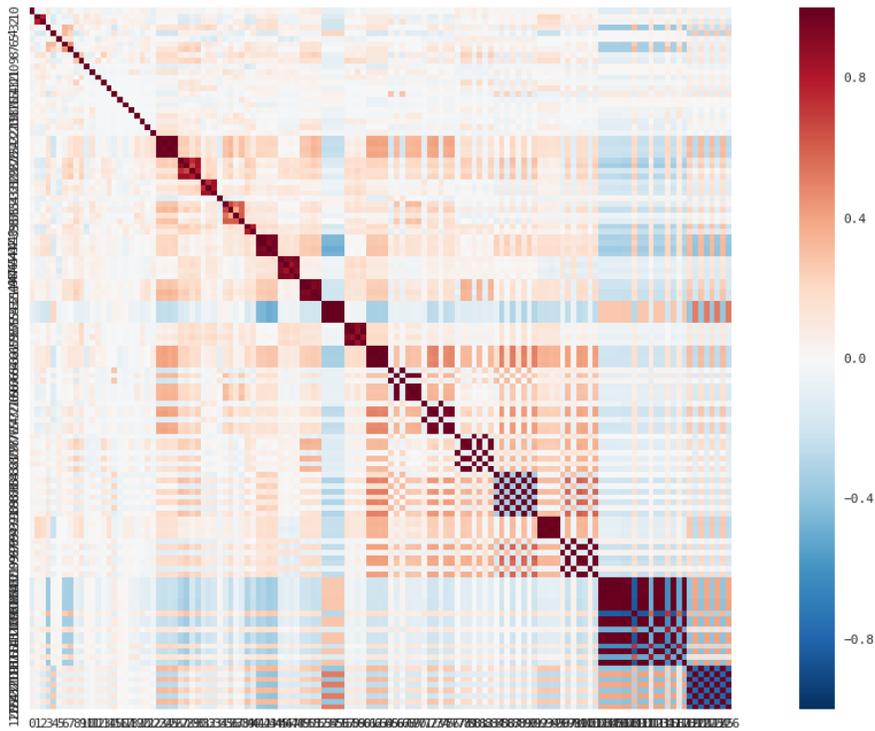
más allá del efecto individual. Otros autores sugieren el uso de umbrales más restrictivos (ej: $p=0.05$), ya que de esta forma se evitan covariables poco significativas para el modelo.

Dado lo anterior, para elegir el umbral p-valor de selección, se lleva a cabo el análisis propuesto por [13], el cual básicamente realiza una búsqueda del umbral p-valor óptimo de inclusión de covariables en base al ranking dado por el test F, en función del desempeño del clasificador SVM para cada umbral. Al imitar el análisis, se comprendió que para los distintos clasificadores seleccionados el umbral óptimo varía (por ejemplo, para el modelo de 24 horas se tienen umbrales óptimos $p=0.3$ y $p=0.25$ para Logit y Naive Bayes respectivamente), por lo cual este se realiza de manera independiente para cada clasificador.

El método propuesto se resume en los siguientes pasos (se realiza lo mismo para cada configuración predictiva: 24, 48 y 72 horas):

1. Se calcula el puntaje F y p-valor para cada uno de los atributos.
2. Se elige un número de posibles umbrales p-valor para la selección de atributos (hacer el corte entre puntajes F altos y bajos). Para los tres modelos se establece $p=0.3$ como el umbral de corte más alto; la razón de esta elección radica en que ampliar más el umbral p-valor implica mayores costos computacionales, además del sobre-ajuste que puede introducir una alta dimensionalidad del set de datos.)
3. Para cada umbral se realiza lo siguiente:
 - Descartar covariables cuyo p-valor esté por debajo del umbral
 - Identificar atributos altamente correlacionados ($abs(\rho)>0.7$), eliminando el que tenga menor puntaje F. La figura 5.1 ilustra la matriz de calor que representa los coeficientes de correlación de Pearson para el corte $p<0.1$ para la base de datos de 24 horas.
 - Realizar validación cruzada de 10 folds para cuatro algoritmos de clasificación: SVM, Logit, Naive Bayes, Artificial Neural Networks. Repetir validación cruzada en 10 iteraciones, y calcular AUC de validación promedio para cada clasificador.
4. Para cada clasificador elegir el umbral de corte con mayor promedio AUC (umbral óptimo).

Figura 5.1: Mapa de calor covariables modelo 24 horas



Fuente: Elaboración Propia

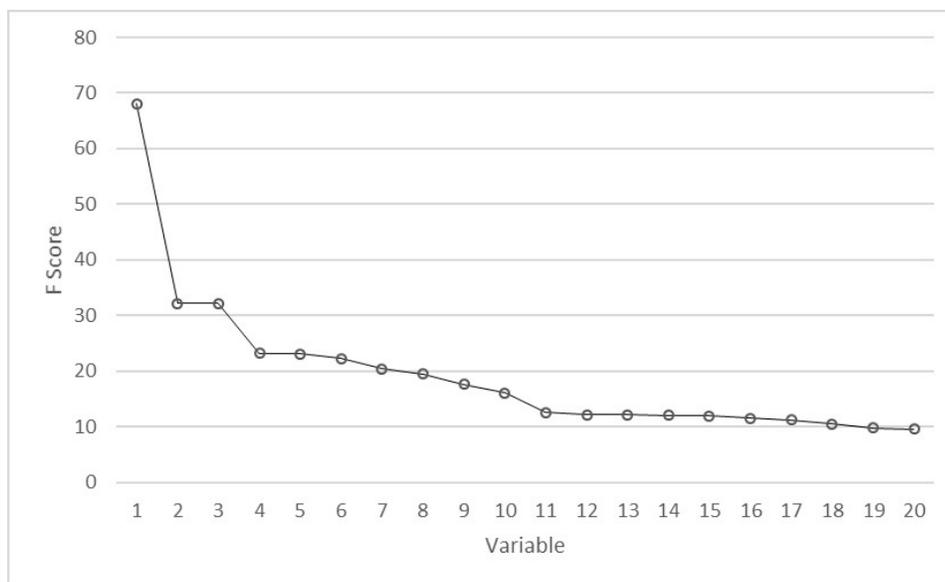
5.2. Modelo 24 horas

5.2.1. Selección de Variables Modelo 24 horas

Siguiendo la metodología que se propone en la sección 5.1, se construye un ranking de variables en función del puntaje F para cada covariable del set de datos de 24 horas, para luego eliminar las variables redundantes según el coeficiente de correlación de Pearson (se eliminan todas las covariables cuyo $\text{abs}(\rho) < 0.7$).

La figura 5.2 corresponde al gráfico que representa el puntaje F para los veinte atributos más relevantes según este criterio, mientras que la tabla 5.2 contiene el índice y descripción de cada variable con respectivos puntaje F y p-valor. El total de variables analizadas y sus correspondientes puntaje F y p-valor se detallan en el Anexo A (solo tomando en cuenta variables cuyo p-valor < 0.3 , que corresponde al umbral más holgado de selección de variables).

Figura 5.2: Puntaje F modelo 24 horas



Fuente: Elaboración Propia

Tabla 5.2: Puntaje F modelo 24 horas

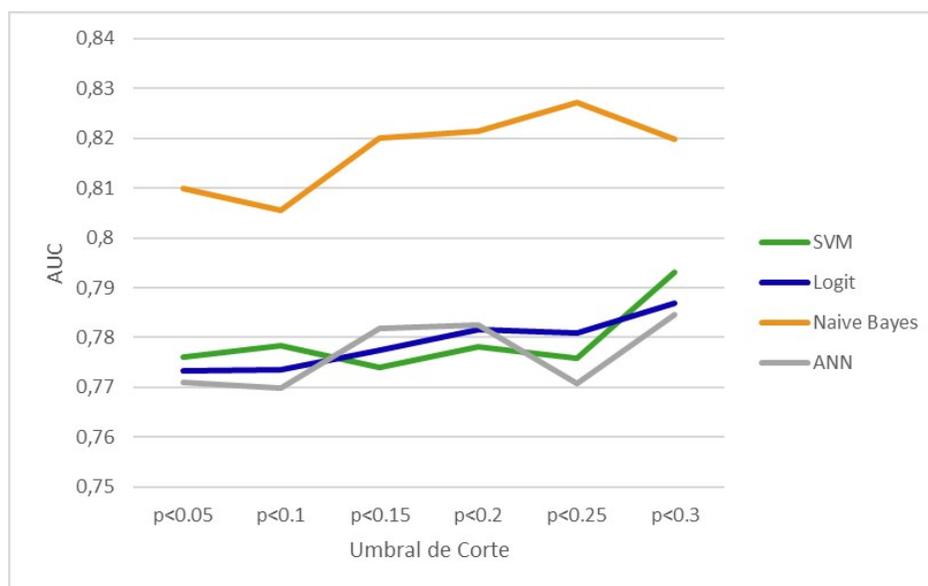
N	Variable	Descripción	F Score	p-valor
1	LAST_AL_HIGH2	Ácido Láctico alto (>4 mmol/L)	68.13	2.24E-11
2	EX21	Coagulopatía	32.17	2.715E-07
3	LAST_PH_LOW1	Ph bajo (<7.25)	32.17	2.71E-07
4	EX14	Enfermedad Hepática	23,15	2.12E-06
5	MEAN_K_HIGH	Potasio alto (>5.1 mEq/L)	23.10	2.18E-06
6	MEAN_LAST_HIGH3	Transaminasas altas (>126/206 U/L)	22.26	3.29E-06
7	LAST_PLT_LOW1	Plaquetas bajas (<80000 K/uL)	20.46	8.05E-06
8	ING_LDH_HIGH	LDH alto (>618 U/L)	19.51	1.29E-05
9	MAX_NA_HIGH	Sodio alto (>145 mEq/L)	17.62	0.00033
10	VM	Ventilación Mecánica	16.11	7.12E-05
11	MIN_ALB_LOW1	Albúmina baja (<2 g/dL)	12.55	0.00044
12	profilaxisgas	Profilaxis gástrico	12.15	0.00054
13	LAST_P_HIGH2	Fosforo alto (>5.5 mg/dL)	12.15	0.00054
14	LOG_HOSP	Días previos hospital (log)	12.08	0.00056
15	LAST_BUN_HIGH2	Nitrógeno Uréico alto (>40 mg/dL))	12.00	0.00059
16	ING_GLC_LOW	Glucosa baja (<75 mg/dL)	11.58	0.00073
17	MIN_WBC_LOW1	Leucocitos bajos (<2000 K/ul)	11.21	0.00089
18	MIN_CRC_LOW1	Acl. Creatinina bajo (<30 ml/min)	10.53	0.00127
19	LAST_CR_HIGH2	Creatinina alta (>2 mg/dL)	9.76	0.00191
20	MIN_SO2_LOW1	Saturación Oxígeno bajo (<60%)	9.58	0.00211

Fuente: Elaboración propia

Luego, se eligen 6 umbrales p-valor de comparación, entre $p=0.05$ (escenario restrictivo), y $p=0.03$ (escenario holgado): $p=0.05$, $p=0.1$, $p=0.15$, $p=0.2$ y $p=0.25$, y $p=0.3$. Para cada uno de estos umbrales se entrena cada uno de los cuatro clasificadores seleccionados (SVM, Logit, Naive Bayes, ANN), y se calcula el valor promedio del área bajo la curva (AUC) producto de la validación cruzada de 10 folds, repetida en 10 iteraciones. Luego para cada clasificador se elige el umbral de corte que demuestra mejor desempeño (número óptimo de variables).

Los resultados de este análisis se resumen en la tabla 5.3, y de manera gráfica mediante la figura 5.3, la cual muestra el resultado del AUC promedio para cada uno de los cuatro algoritmos, para cada umbral incluido en el análisis.

Figura 5.3: Gráfico promedio AUC de validación cruzada para cuatro casificadores, según cada umbral de corte para base de datos 24 horas



Fuente: Elaboración Propia

Tabla 5.3: Resumen resultados promedio AUC de validación cruzada para cuatro casificadores, según cada umbral de corte para base de datos 24 horas

Clasificador	$p \leq 0.05$	$p \leq 0.1$	$p \leq 0.15$	$p \leq 0.2$	$p \leq 0.25$	$p \leq 0.3$	Promedio
SVM	0.776	0.778	0.774	0.778	0.776	0.793	0.779
Logit	0.773	0.774	0.778	0.782	0.781	0.787	0.779
Naive Bayes	0.810	0.805	0.820	0.821	0.827	0.820	0.817
ANN	0.771	0.770	0.782	0.783	0.771	0.785	0.777

Fuente: Elaboración propia

De la figura 5.3 se observa que para todos los umbrales el clasificador NB obtuvo valores AUC notoriamente superiores que los otros tres algoritmos. Por otro lado, para todos los clasificadores se observa una tendencia creciente a medida que se amplía el umbral de inclusión de variables.

Según estos resultados, para Logit y SVM se selecciona $p=0.3$ como umbral de inclusión, abarcando 128 covariables en el modelo. Para NB se selecciona $p=0.25$, el cual representa el máximo local para valores analizados, incluyendo 122 covariables en el modelo predictivo, y, por último, para ANN se selecciona $p=0.15$, el cual, si bien no representa el valor promedio AUC máximo, el cual corresponde al corte $p=0.3$, este es levemente inferior, obteniendo la ventaja de incluir menos variables (82 en vez de 128), reduciendo los costos computacionales, lo cual es relevante dado que este algoritmo es especialmente costoso.

5.2.2. Resultados Entrenamiento Modelo 24 horas

Para cada uno de los clasificadores, se utilizan las variables seleccionadas según el umbral óptimo definido en el análisis realizado en la sección anterior ($p=0.3$ para Logit y SVM, $p=0.25$ para NB y $p=0.15$ para ANN), y se vuelven a entrenar los modelos mediante validación cruzada de 10 folds, esta vez repitiendo 20 veces para cada clasificador. Los resultados se resumen en la tabla 5.4, la cual incluye el valor promedio de cuatro métricas: área bajo la curva (AUC), accuracy, precisión y recall.

Tabla 5.4: Resumen resultados entrenamiento de modelo de 24 horas para cuatro clasificadores

Clasificador	AUC	Accuracy	Presicion	Recall
SVM	0.796	0.753	0.657	0.546
Logit	0.791	0.754	0.665	0.546
Naive Bayes	0.828	0.734	0.771	0.270
ANN	0.777	0.758	0.675	0.536

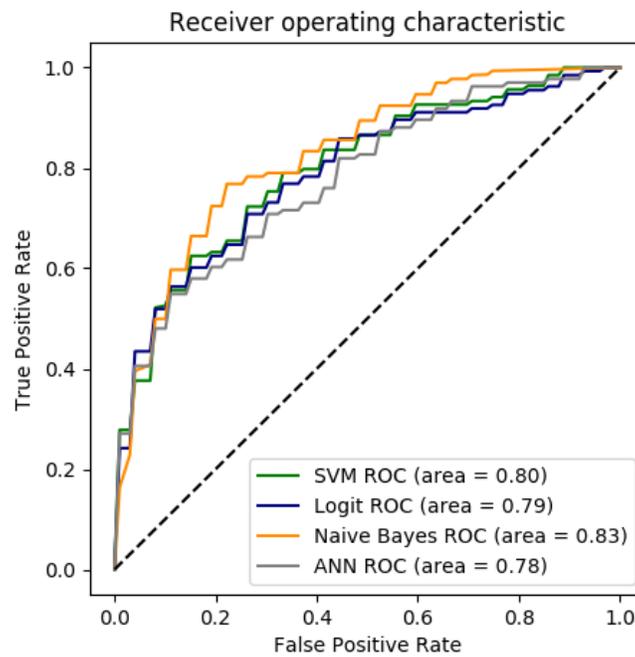
Fuente: Elaboración propia

De la tabla se concluye que, comparando el área bajo la curva (AUC), NB resulta tener el mejor desempeño ($AUC \approx 0.83$) seguido de SVM y Logit ($AUC \approx 0.80$ y $AUC \approx 0.79$ respectivamente), los cuales presentan resultados similares para todas las métricas analizadas, y, por último, con un desempeño levemente menor, ANN con $AUC \approx 0.78$. A pesar de estos resultados, teniendo en cuenta de que AUC es una de las métricas más utilizadas para comparar algoritmos de data mining, este análisis se puede complementar con otras métricas:

Evaluando en base a *accuracy*, es decir, el ratio de casos calificados correctamente en relación al total de observaciones, ANN (0,758) es levemente superior a Logit y SVM (0,754 y 753 respectivamente), mientras que para NB es notoriamente el clasificador con peor desempeño (0,734). Por otro lado, podemos notar que Naive Bayes tiene un puntaje *recall* bajo en relación al puntaje de *precisión* (0,270 y 0,771 respectivamente), lo cual indica que el clasificador tiene un buen desempeño etiquetando casos positivos por sobre el total de etiquetas positivas, no así etiquetando casos positivos por sobre el total de verdaderos positivos, lo que nos indica que este clasificador tiene una baja capacidad para asignar etiquetas positivas, y por lo tanto de predecir que el paciente egresará muerto del servicio. Los otros tres clasificadores demostraron mejor desempeño en cuanto a *recall* (0.536-0.546), lo que indica que etiquetan instancias positivas con mayor frecuencia que Naive Bayes, sin embargo, aún se consideran valores bajos indicando que son restrictivos para clasificar instancias positivas.

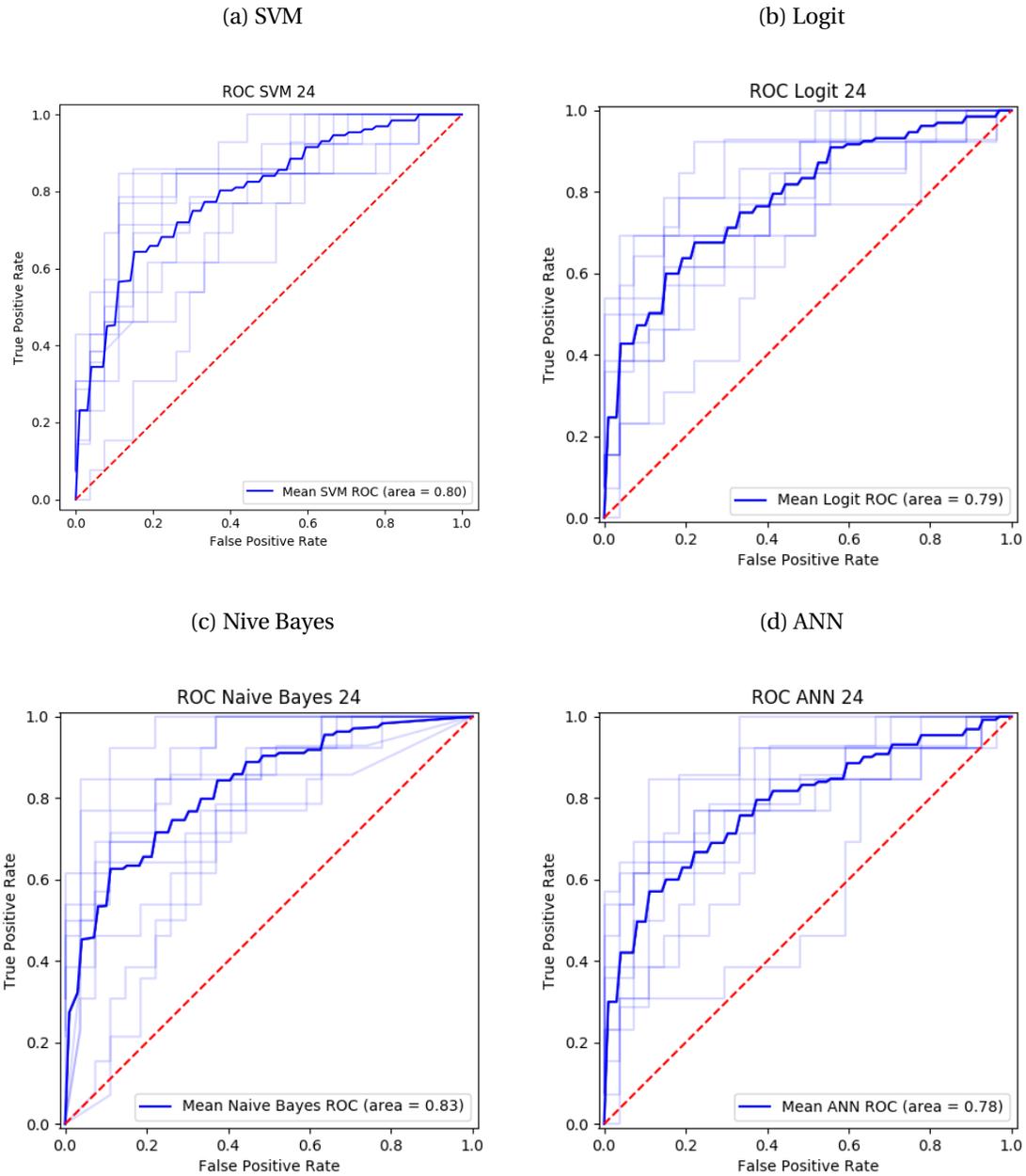
Adicionalmente, la figura 5.5 representa la curva ROC promedio para cada algoritmo de clasificación de mortalidad de 24 horas. Se puede observar que cada punto de la curva corresponde se construye a partir de la curva ROC de cada iteración de la validación cruzada. Por otro lado, la figura 5.4 corresponde a la comparación de la curva ROC promedio de los cuatro algoritmos utilizados en el análisis.

Figura 5.4: Comparación curva ROC promedio para cuatro clasificadores en modelo 24 horas



Fuente: Elaboración Propia

Figura 5.5: Curva ROC para cuatro clasificadores en modelo 24 horas. La curva azul opaca corresponde al promedio de cada curva generada por las iteraciones de la validación cruzada, las cuales se muestran en azul claro



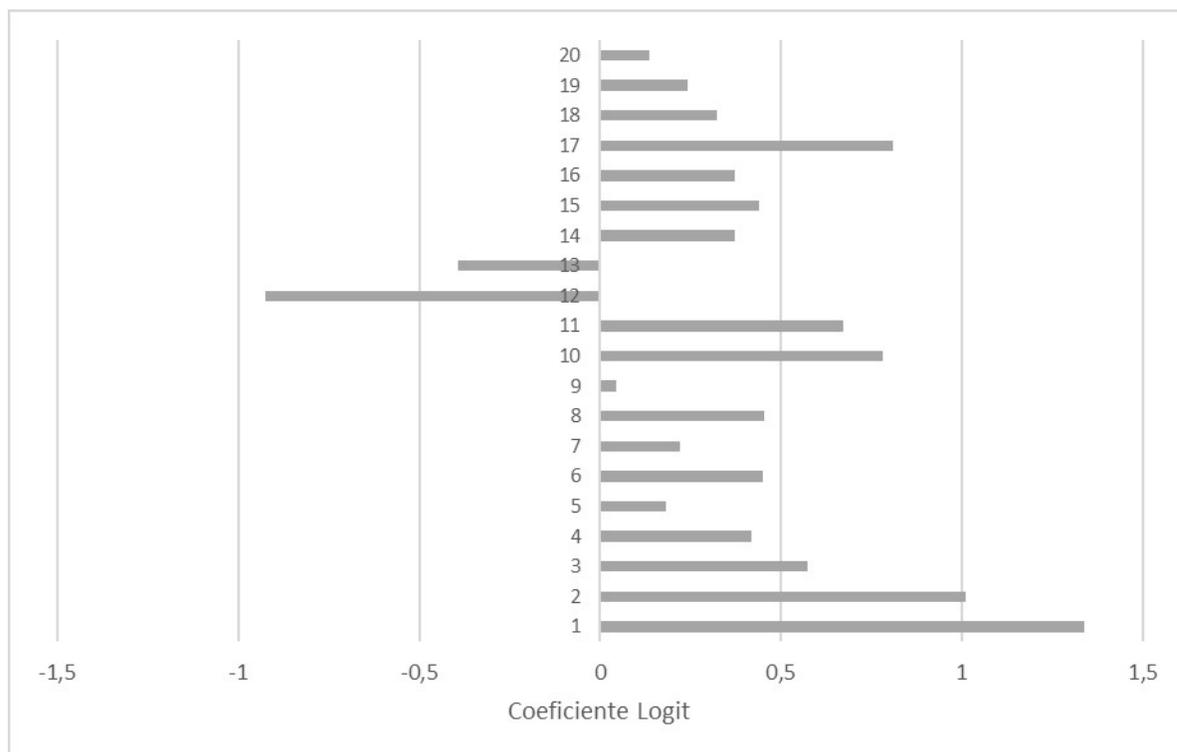
Fuente: Elaboración Propia

5.2.3. Interpretación Modelo 24 horas

Como se menciona en la sección 5.2.1, la regresión logística (Logit) se entrena con las 128 variables más relevantes según el ranking de puntaje F ($p\text{-valor} \leq 0.3$), las cuales se detallan en el Anexo A. La figura 5.6 muestra los coeficientes logit correspondientes a las 20 variables más

relevantes según puntaje F (el detalle con el nombre y descripción de variables se encuentra en la tabla 5.2). El análisis de coeficientes entrega una idea de cuáles son las variables más incidentes para determinar el desenlace del paciente una vez transcurridas las primeras 24 horas de hospitalización en UCI. Los coeficientes logit para las 128 variables incluidas se muestran en el Apéndice B.

Figura 5.6: Coeficientes Logit modelo 24 horas



Fuente: Elaboración Propia

Se destaca la influencia de las siguientes variables protectoras ($\beta < 0$)

- **{12} Profilaxis gástrica** ($\beta = -0,93$): Indica si el paciente consumió uno o más medicamentos de protección gástrica durante las primeras 24 horas de hospitalización.
- **{13} Fósforo alto (último registro)** ($\beta = -0,39$): Indicador de niveles de fósforo muy alto en la sangre ($> 5,5$ mg/dl), en la última medición disponible dentro de las primeras 24 horas de hospitalización. Una de las causas más comunes de la hiperfosfatemia es la presencia de fallo renal agudo, por lo cual ha sido estudiado como factor de riesgo en mortalidad en pacientes críticos por algunos autores [30], concluyendo que estos resultados son controversiales y que se requiere de mayor información y estudio para concluir sobre los efectos del fosfato en mortalidad.

Se destaca la alta influencia de los siguientes factores de riesgo de mortalidad ($\beta > 0,75$):

- **{1} Ácido Láctico alto (último)** ($\beta \approx 1,4$): Indica la presencia de ácido láctico muy alto ($>$

4 mmol/L) en la última medición, lo cual puede ser signo de variadas afecciones, pero se asocia comúnmente a cuadros infecciosos agudos (sepsis). Está ampliamente descrito en la literatura que la hiperlactatemia es un factor de riesgo significativo para predecir mortalidad en pacientes críticos: [47] selecciona 33 estudios que estudian la relación entre niveles de lactato y mortalidad en UCI, concluyendo que la correlación existente es aceptable y significativa para la mayor parte de ellos.

- **{2} Coagulopatía** ($\beta \approx 1,01$): Indica la presencia de alguna coagulopatía como diagnóstico secundario del paciente, como coagulación intravascular diseminada (CID) u otros defectos coagulatorios.
- **{17} Leucocitos bajos (mínimo)** ($\beta \approx 0,81$): Indica niveles muy bajos de glóbulos blancos ($WBC < 2000$), para el valor mínimo registrado durante las primeras 24 horas de hospitalización. Esta ha sido una variable ampliamente estudiada; APACHE II[41] y SAPS II[24] entre otros sistemas de scoring y modelos de mortalidad en UCI, incluyen el conteo de glóbulos blancos (tanto si está por sobre o bajo del rango normal) como factor de riesgo en el puntaje de riesgo del paciente crítico
- **{10} Ventilación Mecánica** ($\beta \approx 0,78$): Indica la presencia de ventilación mecánica invasiva. Se ha documentado que los pacientes que necesitan terapias de soporte vital como lo es VM, son los más graves, y por lo tanto más riesgosos, identificando este como un factor de riesgo no solo en cuanto a mortalidad intra-hospitalaria, sino además en mortalidad a mediano y largo plazo y en la calidad de vida del paciente sobreviviente. Sin embargo, según [19] los resultados del paciente dependen tanto de los factores presentes al comienzo de la ventilación mecánica (configuración de la ventilación), además de la evolución de las complicaciones, y de la cantidad de días/horas en VM, información que sería potencialmente útil para comprender mejor el rol de esta variable en el desenlace de cada paciente.

Se destaca la influencia de los siguientes factores de riesgo de mortalidad ($0,25 < \beta < 0,75$):

- **{11} Albúmina baja (mínimo)** ($\beta \approx 0,67$): Indica niveles de albúmina muy bajos (< 2 gr/dl), en el mínimo valor registrado durante las primeras 24 horas de hospitalización en UCI. El valor de albúmina no es incluido como factor de riesgo en los sistemas de scoring clásicos, sin embargo, existen estudios que concluyen que la hipoalbuminemia está estrechamente asociada con peores desenlaces clínicos[71].
- **{3} Ph bajo (último)** ($\beta \approx 0,57$): Indica un nivel de Ph en la sangre muy bajo ($< 7,25$) en el último registro de las primeras 24 horas de hospitalización en UCI, afección conocida como acidosis metabólica. El nivel de Ph anormal en la sangre, tanto si está por sobre (alcalosis) o bajo (acidosis) los rangos normales, es reconocido como un factor de gravedad en pacientes críticos, incluido por APACHE II[41] en el cálculo del puntaje de riesgo.
- **{8} LDH alto (promedio)** ($\beta \approx 0,46$): Indica un nivel promedio de deshidrogenasa láctica (LDH) muy alto (> 313 u/l) durante 24 horas de hospitalización en UCI.

- **{6} Transaminasas altas (promedio)** ($\beta \approx 0,45$): Indica un nivel promedio de transaminasas GOT (AST) muy alto ($72/118 < \text{AST} \leq 126/206$ u/l (F/M)) durante 24 horas de hospitalización, comúnmente asociado a enfermedades hepáticas, tales como la hepatitis o fallo hepático agudo. Al respecto, no existen estudios que evalúen el impacto individual de la anormalidad en niveles de enzimas hepáticas sobre la mortalidad en pacientes críticos.
- **{15} Nitrógeno Uréico alto (promedio)** ($\beta \approx 0,44$): Indica un nivel promedio de nitrógeno uréico muy alto (> 40 mg/dl) durante 24 horas de hospitalización en UCI, que comúnmente se asocia a insuficiencia cardíaca congestiva o a insuficiencia renal. Este es reconocido como factor de riesgo en mortalidad intra-hospitalaria en pacientes críticos, incluido en SAPS II [24] en el puntaje de gravedad.
- **{4} Enfermedad Hepática** ($\beta \approx 0,42$): Indica la presencia de alguna enfermedad hepática como diagnóstico secundario del paciente, como por ejemplo hepatitis crónica, fallo hepático, cirrosis, entre otras.
- **{14} Días previos hospital (log)** ($\beta \approx 0,37$): Corresponde al logaritmo de la cantidad de días previos que el paciente ha estado hospitalizado antes de ingresar en UCI. Estos resultados sugieren que estadías previas más largas, es decir que el paciente tarda más tiempo en llegar al servicio, se asocia a peores desenlaces.
- **{16} Glucosa baja (ingreso)** ($\beta \approx 0,33$): Indica niveles de glucosa por debajo del rango normal (< 75 mg/dl), en el primer valor registrado durante las primeras 24 horas de hospitalización en UCI, afección conocida como hipoglucemia. El caso contrario, es decir que el paciente ingrese con valores de glucosa superiores a lo normal (hiperglucemia), reconocido por la literatura como factor de riesgo de mortalidad intra-hospitalaria, ocurre con mayor frecuencia siendo muy común en pacientes críticos, incluso en aquellos que no padecen diabetes. Esto ha sido documentado por diversos estudios, destacando el hallazgo de estrechas relaciones entre la implementación de procedimientos de control de glucosa y reducción de mortalidad [45, 22]. En cuando a la hipoglicemia, múltiples estudios reportan relaciones significativas entre hipoglicemia y mortalidad [32, 37, 45, 17], lo que representa evidencia conflictiva ya que la hipoglucemia en algunas ocasiones puede ser causa de la administración excesiva de insulina (control de glucosa).
- **{18} Aclaramiento Creatinina bajo (mínimo)** ($\beta \approx 0,32$): Indica niveles de aclaramiento de creatinina muy bajos ($< 30 \text{ ml/min}$), en el mínimo valor registrado durante las primeras 24 horas de hospitalización en UCI. El aclaramiento de creatinina es una forma de medir cómo están funcionando los riñones, comparando el nivel de creatinina en la sangre y en la orina, con lo cual un nivel muy bajo por lo general es indicador de una insuficiencia renal aguda, afección que se asocia con mayores riesgos de mortalidad en pacientes críticos [1, 57].

Se destaca una influencia menor de los siguientes factores de riesgo de mortalidad ($\beta < 0,25$)

- **{19} Creatinina alta (último)** ($\beta \approx 0,24$): Indica niveles de creatinina muy altos ($> 2\text{ mg/dl}$), en el último valor registrado durante las primeras 24 horas de hospitalización en UCI. El nivel de creatinina en la sangre es el marcador comúnmente más usado para evaluar el funcionamiento del sistema renal, asociando niveles superiores a insuficiencia renal, siendo indicador de gravedad y mayor riesgo de muerte intra-hospitalaria en sistemas de scoring clásicos que se utilizan en UCI, APACHE II[41], SOFA[70], SAPS II[24] y MPM II[50].
- **{7} Plaquetas bajas (mínimo)** ($\beta \approx 0,22$): Indica un conteo de plaquetas muy inferior al rango normal ($< 80000\text{ K/UL}$), en el mínimo valor registrado durante las primeras 24 horas de hospitalización en UCI. Está documentado que niveles bajos de plaquetas es una anomalía común en pacientes UCI, y que además representa un factor de riesgo independiente de mortalidad[69].
- **{5} Potasio alto (promedio)** ($\beta \approx 0,18$): Indica un nivel de potasio por sobre el rango normal ($K > 5,1$) para el promedio de todas las mediciones disponibles durante 24 horas de hospitalización en UCI, afección conocida como hipercalemia. Niveles altos potasio son reconocidas como factores predictivos de mortalidad tanto por sistemas de scoring clásicos como APACHE II [41] y SAPS II [24], como por otros estudios [68, 33].
- **{20} Saturación Oxígeno baja (mínimo)** ($\beta \approx 0,14$): Indica saturación de oxígeno en la sangre muy baja ($< 60\%$), en el mínimo valor registrado durante las primeras 24 horas de hospitalización en UCI, es decir, que el paciente esté hipoxémico. Los niveles de oxígeno en la sangre (SatO₂, PO₂, FiO₂) han sido identificados en distintos estudios[72] y modelos de riesgo para predecir mortalidad en pacientes críticos: APACHE II[41], SAPS II[24], MPM II[50] y SOFA[70].
- **{9} Sodio alto (máximo)** ($\beta \approx 0,046$): Indica un nivel de sodio en la sangre por sobre el rango normal ($NA > 145\text{ mEq/L}$) para el valor máximo obtenido durante 24 horas de hospitalización en UCI. Se ha documentado que tanto la hiponatremia (sodio bajo), como la hipernatremia (sodio alto), son ambos factores de riesgo independientes en la predicción de mortalidad en pacientes críticos [?], siendo incluido como factor de riesgo en los sistemas de scoring APACHE II[41] y SAPS II[24]. Respecto de los resultados aquí obtenidos, se concluye que para obtener más información, en trabajos futuros se sugiere utilizar más rangos de valores.

5.3. Modelo 48 horas

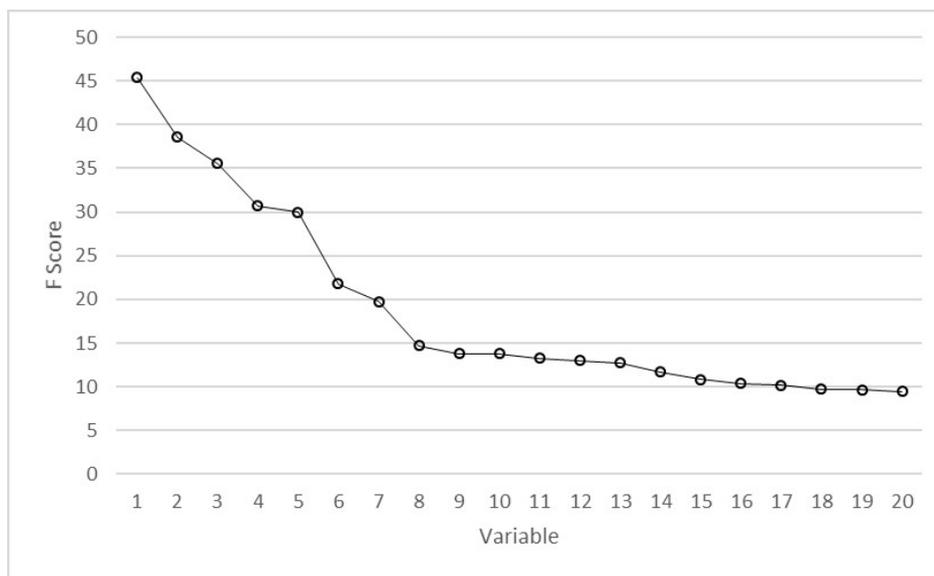
5.3.1. Selección de Variables Modelo 48 horas

Siguiendo la metodología que se propone en la sección 5.1, se construye un ranking de variables en función del puntaje F para cada covariable del set de datos de 48 horas, que incluye

364 observaciones, para luego eliminar las variables redundantes según el coeficiente de correlación de Pearson (se eliminan todas las covariables cuyo $\text{abs}(\rho) < 0.7$). La figura 5.7 corresponde al gráfico que representa el puntaje F para los veinte atributos más relevantes según este criterio, mientras que la tabla 5.7 contiene el índice y descripción de cada variable con respectivo puntaje.

El total de variables analizadas y sus correspondientes puntaje F y p-valor se detallan en el Anexo A (solo tomando en cuenta variables cuyo p-valor ≤ 0.3 , que corresponde al umbral más holgado de selección de variables).

Figura 5.7: Puntaje F modelo 48 horas



Fuente: Elaboración Propia

Tabla 5.5: Puntaje F modelo 48 horas

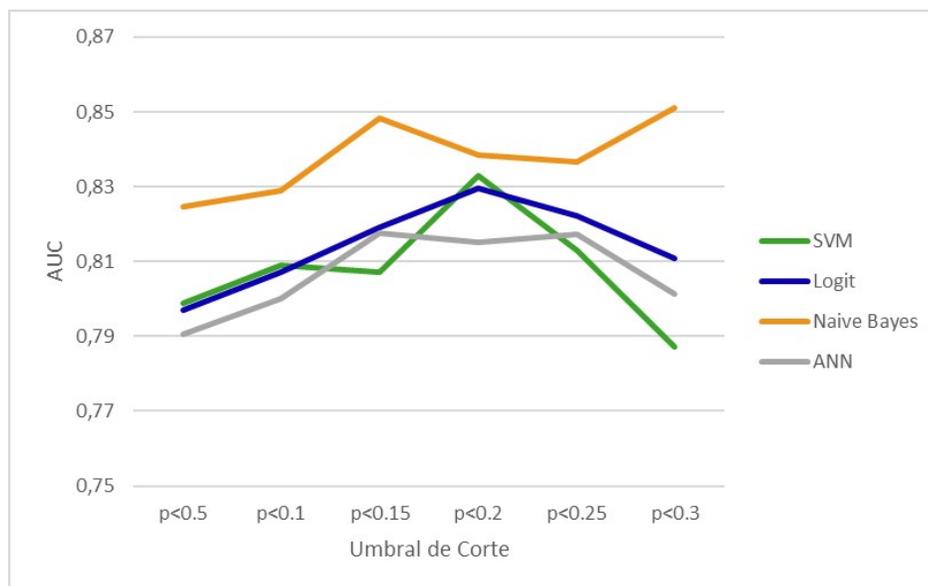
N	Variable	Descripción	F Score	p-valor
1	LAST_PH_LOW1	Ph bajo (<7.25)	45.37	6.43E-09
2	EX21	Coagulopatía	38.54	1.47E-08
3	LAST_AL_HIGH2	Ácido Láctico alto (>4 mmol/L)	35.54	5.95E-08
4	LAST_AST_HIGH3	Transaminasas altas (>126/206 U/L)	30.71	5.77E-07
5	LAST_PLT_LOW1	Plaquetas bajas (<80000 K/uL)	30.00	8.09E-07
6	EX14	Enfermedad Hepática	21.76	4.35E-06
7	LAST_LDH_HIGH	LDH alto (>618 U/L)	19.77	1.17E-05
8	LAST_CRC_LOW1	Acl. creatinina bajo (<30 ml/min)	14.72	0.00015
9	LAST_K_HIGH	Potasio alto (>5.1 mEq/L)	13.77	0.00024
10	LAST_CR_HIGH2	Creatinina alta (>2 mg/dL)	13.76	0.00024
11	LOG_HOSP	Días previos hospital (log)	13.26	0.00031
12	LAST_BUN_HIGH2	Nitrógeno Uréico alto (>40 mg/dL)	13.03	0.00035
13	MEAN_P_HIGH2	Fosforo alto (>5.5 mg/dL)	12.75	0.00040
14	LAST_BLR_HIGH2	Bilirrubina alta (> 2 mg/dL)	11.68	0.000711
15	VM	Ventilación Mecánica	10.83	0.00109
16	MIN_WBC_LOW1	Leucocitos bajos (<2000 K/uL)	10.35	0.00141
17	LAST_AL_HIGH1	Ácido Láctico alto (2<AL<4 mmol/L)	10.20	0.00153
18	LAST_PCT_HIGH3	Procalcitonina alta (>4 ng/L)	9.70	0.00199
19	MIN_WBC_HIGH1	Leucocitos altos (10k<WBC<15k K/uL)	9.67	0.00202
20	LAST_PCT_MISS	Procalcitonina ausente	9.43	0.00229

Fuente: Elaboración propia

Luego, se eligen 6 umbrales p-valor de comparación, entre $p=0.05$ (escenario restrictivo), y $p=0.03$ (escenario holgado): $p=0.05$, $p=0.1$, $p=0.15$, $p=0.2$ y $p=0.25$, y $p=0.3$. Para cada uno de estos umbrales se entrena cada uno de los cuatro clasificadores seleccionados (SVM, Logit, Naive Bayes, ANN), y se calcula el valor promedio del área bajo la curva (AUC) producto de la validación cruzada de 10 folds, repetida en 10 iteraciones. Luego para cada clasificador se elige el umbral de corte que demuestra mejor desempeño (número óptimo de variables).

Los resultados de este análisis se resumen en la tabla 5.6, y de manera gráfica mediante la figura 5.8, la cual muestra el resultado del AUC promedio para cada uno de los cuatro algoritmos, para cada umbral incluido en el análisis.

Figura 5.8: Gráfico promedio AUC de validación cruzada para cuatro clasificadores, según cada umbral de corte para base de datos 48 horas



Fuente: Elaboración Propia

Tabla 5.6: Resumen resultados promedio AUC de validación cruzada para cuatro clasificadores, según cada umbral de corte para base de datos 48 horas

Clasificador	p ≤ 0.05	p ≤ 0.1	p ≤ 0.15	p ≤ 0.2	p ≤ 0.25	p ≤ 0.3	Promedio
SVM	0.799	0.809	0.807	0.833	0.813	0.787	0.808
Logit	0.797	0.807	0.829	0.830	0.822	0.811	0.814
Naive Bayes	0.825	0.829	0.848	0.838	0.837	0.851	0.838
ANN	0.791	0.800	0.818	0.815	0.817	0.801	0.807

Fuente: Elaboración propia

De la figura 5.8, se observa que al igual que el modelo de 24 horas, para todos los umbrales el clasificador NB posee valores AUC notoriamente superiores que los otros tres algoritmos. Para Logit y SVM se observa una tendencia creciente hasta el corte $p=0.2$ ($AUC \approx 0.83$ en ambos casos), punto desde el cual el desempeño del modelo comienza a decrecer, razón por la cual se elige este punto como umbral de corte para estos algoritmos, seleccionando 114 covariables para entrenar el modelo.

En el caso de NB y ANN, se observa una tendencia creciente hasta el punto $p=0.15$ ($AUC \approx 0.85$ y $AUC \approx 0.82$ respectivamente), punto desde el cual decrece el desempeño para volver a subir en $p=0.25$ para ANN y $p=0.3$ para NB. Para ambos algoritmos se selecciona $p=0.15$ como umbral de corte por el beneficio de construir modelos menos costosos, abarcando 99 covariables para el entrenamiento del modelo.

5.3.2. Resultados Entrenamiento Modelo 48 horas

Para cada uno de los clasificadores, se utilizan las variables seleccionadas según el umbral óptimo definido en el análisis realizado en la sección anterior ($p=0.2$ para Logit y SVM, $p=0.15$ para NB y ANN), y se vuelven a entrenar los modelos mediante validación cruzada de 10 folds, esta vez repitiendo 20 veces para cada clasificador. Los resultados se resumen en la tabla 5.5, la cual incluye el valor promedio de cuatro métricas: área bajo la curva (AUC), accuracy, precisión y recall.

Tabla 5.7: Resumen resultados entrenamiento de modelo de 48 horas para cuatro clasificadores

Clasificador	AUC	Accuracy	Presicion	Recall
SVM	0.828	0.799	0.730	0.575
Logit	0.825	0.797	0.737	0.575
Naive Bayes	0.849	0.744	0.828	0.203
ANN	0.816	0.789	0.714	0.541

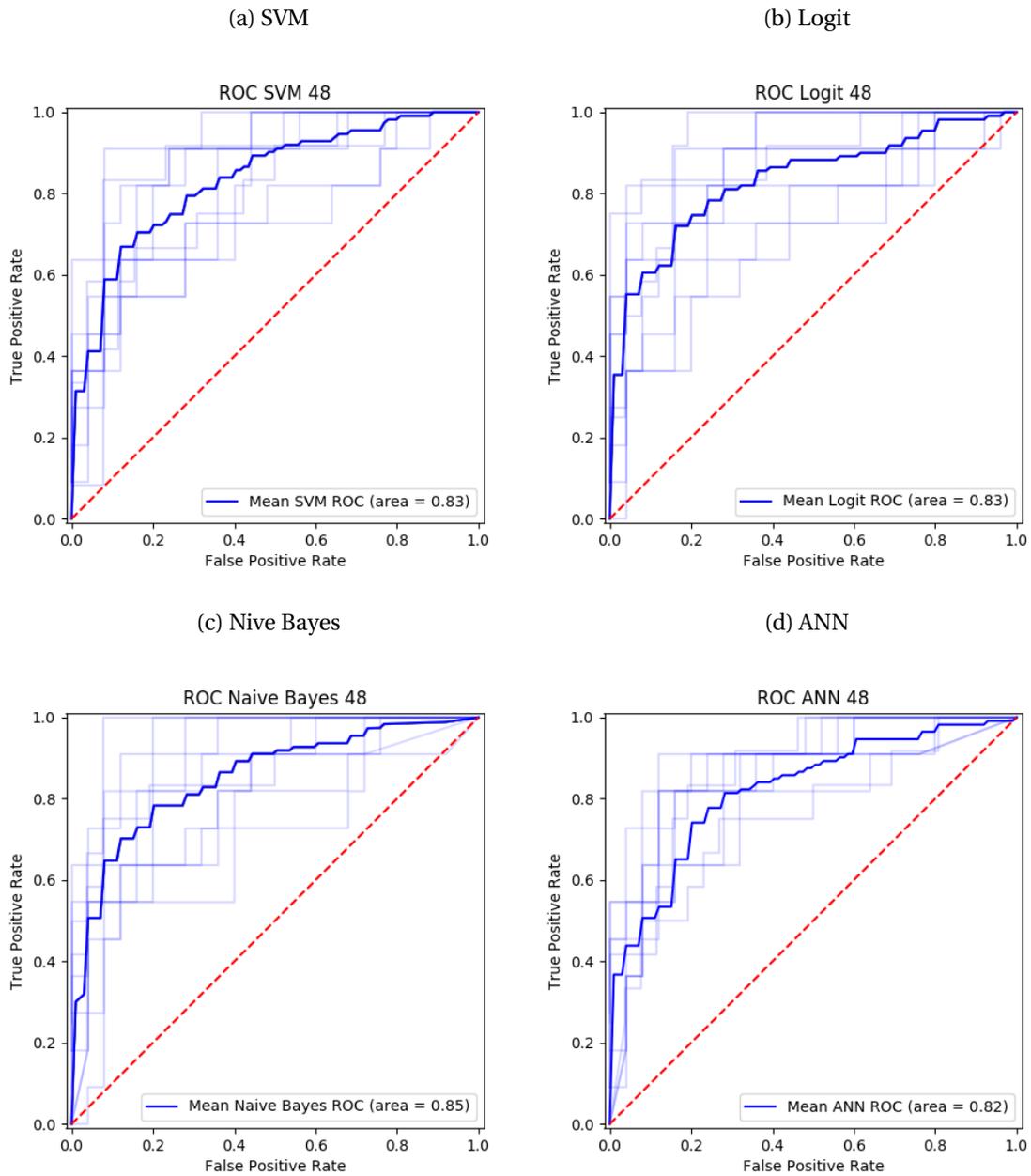
Fuente: Elaboración propia

Al igual que el modelo de 24 horas, NB representa el modelo con mejor desempeño comparando el área bajo la curva ROC ($AUC \approx 0.85$), seguido de SVM y Logit ($AUC \approx 0.83$) y luego, de ANN con un desempeño levemente menor ($AUC \approx 0.82$).

Mirando las otras métricas, se observa que SVM y Logit corresponden a los modelos con mejor *accuracy*, es decir el porcentaje de instancias etiquetadas correctamente ($acc \approx 0.8$), seguido por ANN ($acc \approx 0.79$) y luego NB con un desempeño sustancialmente menor ($acc \approx 0.74$). Al igual que para el modelo de 24 horas, se confirma que NB posee una precisión y AUC bastante superiores, pero un *recall* inferior al resto de los clasificadores, dejando en evidencia que este algoritmo se desempeña bien en el ajuste de etiquetas positivas (casos positivos de mortalidad asignados correctamente), pero que es inferior en la cantidad de etiquetas positivas que asigna, es decir, que subestima el número de casos positivos.

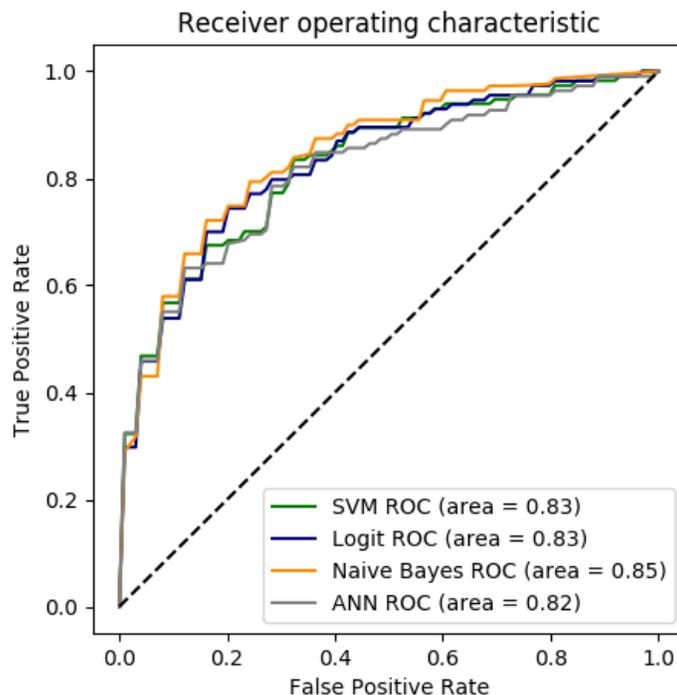
Adicionalmente, se realiza un análisis gráfico del desempeño del modelo mediante la figura 5.9, la cual representa la curva ROC promedio para cada algoritmo de clasificación de mortalidad de 48 horas. Por otro lado, la figura 5.10 corresponde a la comparación de la curva ROC promedio de los cuatro algoritmos utilizados en el análisis.

Figura 5.9: Curva ROC para cuatro clasificadores en modelo 48 horas. La curva azul opaca corresponde al promedio de cada curva generada por las iteraciones de la validación cruzada, las cuales se muestran en azul claro



Fuente: Elaboración Propia

Figura 5.10: Comparación curva ROC promedio para cuatro clasificadores en modelo 48 horas



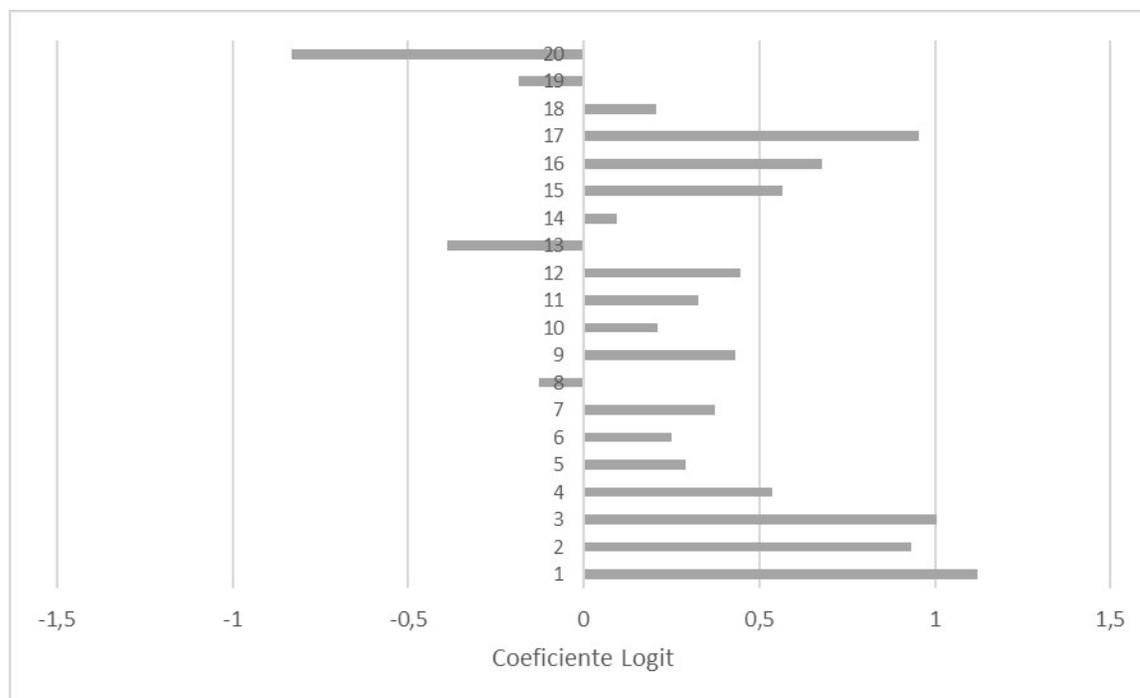
Fuente: Elaboración Propia

5.3.3. Interpretación Modelo 48 horas

Como se menciona en el capítulo 5.3.1, después de realizar el análisis para encontrar el umbral óptimo de selección de variables, el clasificador Logit se entrena con las 114 covariables más relevantes según el ranking de puntaje F ($p\text{-valor} \leq 0.2$), las cuales se detallan en el Apéndice A.

La figura 5.11 muestra los coeficientes Logit correspondientes a las 20 variables más relevantes según puntaje F (el detalle con el nombre y descripción de variables se encuentra en la tabla 5.5). El análisis de estos coeficientes entrega una idea de cuáles son las variables más incidentes para determinar el desenlace del paciente una vez transcurridas las primeras 48 horas de hospitalización en UCI. Los coeficientes logit del total de 114 variables incluidas en el modelo se muestran en el Apéndice B.

Figura 5.11: Coeficientes Logit para 20 variables más relevantes, modelo 48 horas



Fuente: Elaboración Propia

De las variables seleccionadas, se destaca la influencia las siguientes variables protectoras ($\beta < 0$):

- **{20} Procalcitonina Ausente** ($\beta \approx 0,83$): Indica que al paciente no se le realiza ningún examen de procalcitonina durante las primeras 48 horas de hospitalización. Esto tiene sentido ya que este examen se ordena como marcador diagnóstico en infecciones y sepsis, por lo que intuitivamente, el hecho de que no se realice la orden puede ser un indicador de ausencia de síntomas relacionados a cuadros infecciosos y por lo tanto, que el paciente sea menos riesgoso.
- **{13} Fósforo alto (último)** ($\beta \approx -0,39$): Indicador de niveles de fósforo muy alto en la sangre ($> 5,5$ mg/dl, en la última medición disponible dentro de las primeras 24 horas de hospitalización.(para más detalles ver descripción sección 5.1.3)
- **{19} Leucocitos altos (mínimo)** ($\beta \approx -0,18$): Indica niveles moderadamente elevados de glóbulos blancos por sobre el rango normal ($10000 < WBC \leq 15000$), en el mínimo valor registrado durante las primeras 48 horas de hospitalización en UCI. Como se describe en la sección 5.1.3, niveles de leucocitos, tanto si están por sobre o son superiores a los rangos normales, es un factor de riesgo ampliamente documentado, por lo cual estos resultados se consideran poco concluyentes.
- **{8} Aclaramiento creatinina bajo** ($\beta \approx -0,13$): Indica niveles de aclaramiento de creatini-

na muy bajos ($< 30\text{ml}/\text{min}$), en el mínimo valor registrado durante las primeras 24 horas de hospitalización en UCI. Como se describe en la sección 5.1.3, niveles bajos aclaramiento de creatinina son indicador de fallo renal agudo, representando un factor de riesgo para mortalidad, que, sumado a los resultados de esta variable para el modelo de 24 horas, se concluye que este resultado es contradictorio y poco concluyente.

Se destaca la alta influencia como factor de riesgo de mortalidad de las siguientes variables ($\beta > 0,75$):

- **{1} Ph bajo** ($\beta \approx 1,12$): Indica un nivel de Ph en la sangre muy bajo ($< 7,25$) en el último registro de las primeras 24 horas de hospitalización en UCI (para más detalles ver descripción sección 5.1.3)
- **{3}{17} Ácido Láctico alto** ($\beta \approx 0,93$ / $\beta \approx 0,96$): Indica la presencia de ácido láctico alto ($2,1 < \text{AL} \leq 4$) y muy alto ($\text{AL} > 4$) respectivamente (para más detalles ver descripción en sección 5.1.3)
- **{2} Coagulopatía** ($\beta \approx 0,93$): Indica la presencia de alguna coagulopatía, como coagulación intravascular diseminada (CID) u otros defectos coagulatorios.

Se destaca una influencia moderada como factor de riesgo de mortalidad ($0,25 < \beta < 0,75$) de las siguientes variables:

- **{16} Leucocitos bajos** ($\beta \approx 0,68$): Indica niveles muy bajos de glóbulos blancos (< 2000), para el valor mínimo registrado durante las primeras 48 horas de hospitalización (para más detalles ver descripción en sección 5.1.3)
- **{15} Ventilación Mecánica** ($\beta \approx 0,57$): Indica la presencia de ventilación mecánica invasiva (para más detalles ver descripción en sección 5.1.3).
- **{4} Transaminasas altas (último)** ($\beta \approx 0,54$): Indica un nivel promedio de transaminasas (AST) muy alto ($> 126/206$ U/L (F/M)) durante 48 horas de hospitalización (para más detalles ver descripción en sección 5.1.3).
- **{12} Nitrógeno Uréico alto (último)** ($\beta \approx 0,45$): Indica un nivel promedio de nitrógeno uréico muy alto (> 40 mg/dL) (para más detalles ver descripción en sección 5.1.3).
- **{9} Potasio alto (último)** ($\beta \approx 0,43$): Indica un nivel de potasio por sobre el rango normal ($> 5,1$ mEq/L) en la última medición disponible durante 48 horas de hospitalización en UCI (para más detalles ver descripción en sección 5.1.2).
- **{7} LDH alta (último)** ($\beta \approx 0,37$): Indica un nivel de deshidrogenasa láctica (LDH) muy alto (> 618 U/L) durante 48 horas de hospitalización en UCI.

- **{11} Días previos hospital (log)** ($\beta \approx 0,33$): Corresponde al logaritmo de la cantidad de días previos que el paciente ha estado hospitalizado antes de ingresar a UCI. Estos resultados sugieren que estadías previas más largas, es decir que el paciente tarda más tiempo en llegar al servicio, se asocia a peores desenlaces.
- **{5} Plaquetas bajas (último)** ($\beta \approx 0,29$): Indica niveles de conteo de plaquetas por debajo del rango normal (<150.000 K/UI), en el último valor registrado durante las primeras 48 horas de hospitalización en UCI (para más detalles ver descripción en sección 5.1.3).
- **{6} Enfermedad hepática** ($\beta \approx 0,25$): Indica la presencia de alguna enfermedad hepática como diagnóstico secundario del paciente, como por ejemplo hepatitis crónica, fallo hepático, cirrosis, entre otras.

Se destaca una influencia menor de los siguientes factores de riesgo de mortalidad ($\beta < 0,25$)

- **{10} Creatinina alta (último)** ($\beta \approx 0,21$): Indica niveles de creatinina por sobre el rango normal ($> 2mg/dl$), en el último valor registrado durante las primeras 48 horas de hospitalización en UCI. (para más detalles ver descripción en sección 5.1.3).
- **{18} Procalcitonina alta (último)** ($\beta \approx 0,21$): Indica niveles de procalcitonina muy altos ($> 4ng/L$), en el último valor registrado durante las primeras 48 horas de hospitalización en UCI, utilizado como marcador de infección en UCI, y documentado por algunos estudios como indicador independiente de riesgo y gravedad en pacientes críticos [76, 39].
- **{14} Bilirrubina alta (último)** ($\beta \approx 0,09$): Indica niveles de bilirrubina total muy altos ($> 2mg/dl$), en el último valor registrado durante las primeras 48 horas de hospitalización en UCI, lo cual es marcador de presencia de enfermedades hepáticas. Tanto SAPS II[24] como SOFA[70] utilizan este indicador para definir el estado de gravedad del paciente, asociando niveles altos de bilirrubina con peores resultados.

5.4. Modelo 72 horas

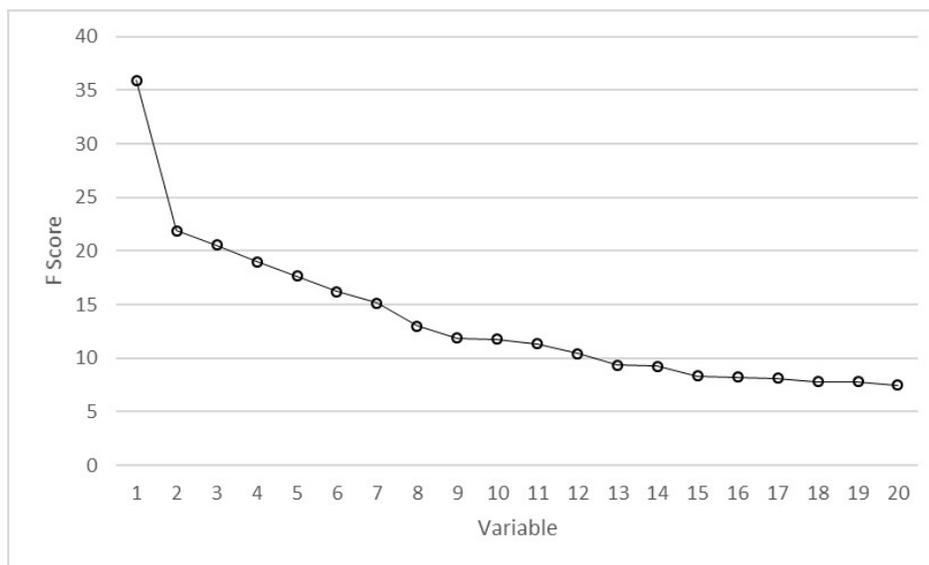
5.4.1. Selección de Variables Modelo 72 horas

Siguiendo la metodología que se propone en la sección 5.1, se construye un ranking de variables en función del puntaje F para cada covariable del set de datos de 72 horas, para luego eliminar las variables redundantes según el coeficiente de correlación de Pearson (se eliminan todas las covariables cuyo $abs(p) < 0,7$). La figura 5.12 corresponde al gráfico que representa el puntaje F para los veinte atributos más relevantes según este criterio, mientras que la tabla 5.8 contiene el índice y descripción de cada variable con respectivo puntaje.

Para el total de variables analizadas y sus correspondientes puntaje F y p-valor se detallan en el Anexo A (solo tomando en cuenta variables cuyo $p\text{-valor} \leq 0,25$, que corresponde al umbral

más holgado de selección de variables).

Figura 5.12: Puntaje F modelo 72 horas



Fuente: Elaboración Propia

Tabla 5.8: Puntaje F modelo 72 horas

N	Variable	Descripción	F Score	p-valor
1	EX21	Coagulopatía	35,8847	5,746E-08
2	MEAN_PLT_LOW1	Plaquetas bajas (<80000 K/uL)	21,885	4,314E-06
3	MEAN_PH_LOW1	Ph bajo (<7.25)	20,5697	8,201E-06
4	MAX_LDH_HIGH	LDH alto (618 U/L)	18,982	1,791E-05
5	LAST_AL_HIGH2	Ácido Láctico alto (>4 mmol/L)	17,6377	3,488E-05
6	MAX_AST_HIGH3	Transaminasas altas (>126/206 U/L)	16,2038	7,14E-05
7	EX14	Enfermedad Hepática	15,1634	0,0001
8	LAST_BLR_HIGH2	Bilirrubina alta (>2 mg/dL)	13,0051	0,0004
9	LOG_HOSP	Días previos hospital (log)	11,895	0,0006
10	LAST_AL_HIGH1	Ácido Láctico alto (2<AL<4 mmol/L)	11,7715	0,0007
11	MIN_HGB_LOW1	Hemoglobina baja (<7 gr/dL)	11,3553	0,0008
12	EX5	Enfermedad Vascular Periférica	10,4398	0,0014
13	C16	Cáncer Estómago	9,368	0,0024
14	MIN_SO2_LOW1	Saturación Oxígeno bajo (<60 %)	9,2316	0,0026
15	MEAN_CR_HIGH2	Creatinina alta (>2 mg/dL)	8,354	0,0041
16	MAX_PH_HIGH2	Ph alto (>7.5)	8,2246	0,0044
17	MAX_UA_LOW	Ácido Úrico bajo (<2.5/3.5 mg/dL)	8,1106	0,0047
18	MEAN_K_HIGH	Potasio alto (>5.1 mEq/L)	7,8337	0,0054
19	LAST_COHG_HIGH	Carboxihemoglobina alta (>2.5 %)	7,8337	0,0054
20	MIN_AL_LOW	Ácido Láctico bajo (<0.7 mmol/L)	7,4636	0,0067

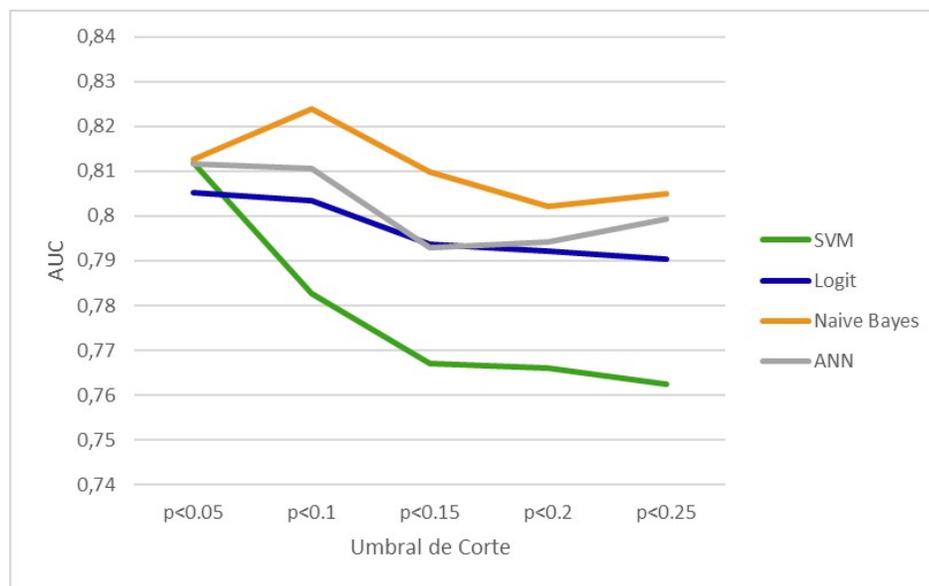
Fuente: Elaboración propia

Luego, se eligen 5 umbrales p-valor de comparación, entre $p=0.05$ (escenario restrictivo), y $p=0.25$ (escenario holgado): $p<0.05$, $p<0.1$, $p<0.15$, $p<0.2$ y $p<0.25$. La razón de esta elección es que, en el caso de todos los clasificadores, el desempeño baja sustancialmente para escenarios más holgados que $p=0.1$, no siendo necesaria la extensión del análisis.

Luego, para cada umbral de corte se entrena cada uno de los cuatro clasificadores seleccionados (SVM, Logit, Naive Bayes y ANN), y calcula el valor promedio del área bajo la curva (AUC) producto de la validación cruzada de 10 folds, repetida en 10 iteraciones. Para cada clasificador se selecciona el umbral de corte que otorga mejores resultados (número óptimo de variables).

Los resultados de este análisis se resumen en la tabla 5.9, y de manera gráfica mediante la figura 5.13, la cual muestra el resultado del AUC promedio para cada uno de los cuatro algoritmos, para cada umbral incluido en el análisis.

Figura 5.13: Gráfico promedio AUC de validación cruzada para cuatro clasificadores, según cada umbral de corte para base de datos 72 horas



Fuente: Elaboración Propia

Tabla 5.9: Resumen resultados promedio AUC de validación cruzada para cuatro clasificadores, según cada umbral de corte para base de datos 72 horas

Clasificador	$p\leq 0.05$	$p\leq 0.1$	$p\leq 0.15$	$p\leq 0.2$	$p\leq 0.25$	Promedio
SVM	0.812	0.783	0.767	0.766	0.762	0.778
Logit	0.805	0.804	0.794	0.792	0.790	0.797
Naive Bayes	0.813	0.824	0.810	0.802	0.805	0.811
ANN	0.812	0.811	0.793	0.794	0.799	0.802

Fuente: Elaboración propia

En este caso NB sigue siendo superior al resto de los clasificadores, pero, a diferencia de los

modelos de 24 y 48 horas, la diferencia en el desempeño es menor, siendo incluso igual para el punto de corte más bajo ($p=0.05$). Por otro lado, para SVM, Logit y ANN, se observa una tendencia decreciente en el AUC promedio a medida que se amplía el umbral de inclusión de variables, por lo cual se selecciona $p=0.05$ como umbral de corte para estos algoritmos, abarcando 47 covariables. En el caso de NB se selecciona $p=0.1$ (65 covariables) como umbral de corte dado que este representa el máximo como promedio AUC, punto desde el cual el desempeño decrece sustancialmente.

5.4.2. Resultados Entrenamiento Modelo 72 horas

En esta sección se reportan los resultados del entrenamiento del modelo de 72 horas para cada uno de los clasificadores. Como técnica de entrenamiento se aplica validación cruzada de 10 folds, repitiendo en 20 iteraciones para cada modelo. Los resultados se resumen en la tabla 5.10, en la cual se incluye el valor promedio de cuatro métricas: el área bajo la curva ROC (AUC), accuracy, precisión y recall.

Tabla 5.10: Resumen resultados entrenamiento de modelo de 72 horas para cuatro clasificadores

Clasificador	AUC	Accuracy	Presicion	Recall
SVM	0.808	0.795	0.714	0.486
Logit	0.805	0.800	0.726	0.486
Naive Bayes	0.823	0.794	0.835	0.341
ANN	0.805	0.779	0.652	0.495

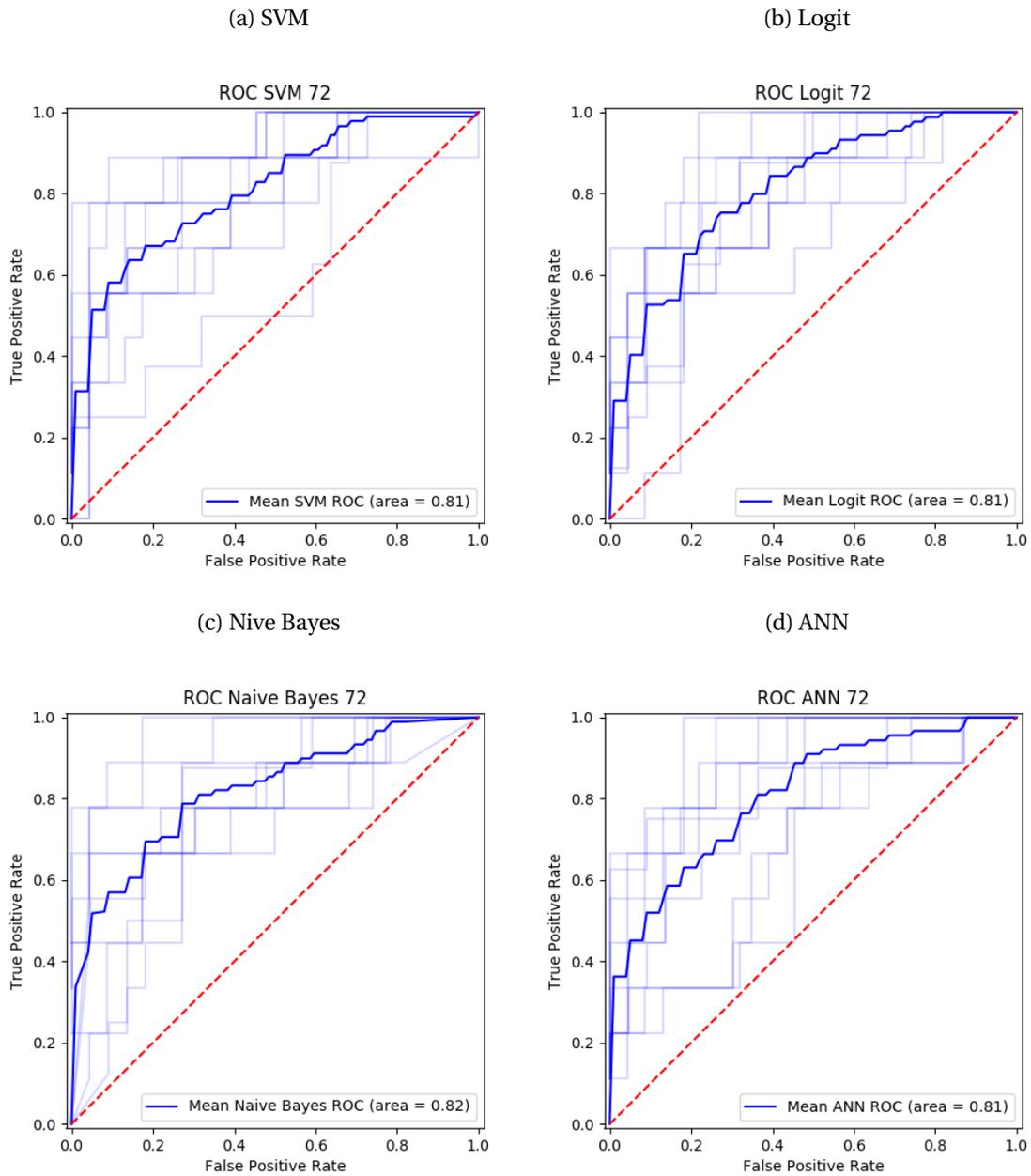
Fuente: Elaboración propia

Al igual que los modelos de 24 y 48 horas, Naive Bayes representa el modelo con mejor desempeño comparando el área bajo la curva ROC ($AUC \approx 0.82$), seguido de SVM y Logit y ANN que poseen resultados similares ($AUC \approx 0.81$). Como se menciona en la sección anterior, para este modelo la brecha entre NB y el resto de los algoritmos se acorta, reduciendo así la ventaja que tiene usar este por sobre los otros tres.

En cuanto a *accuracy*, es decir la proporción de observaciones bien clasificadas, SVM, Logit y NB obtienen puntajes similares ($acc \approx 0.8$), mientras que ANN obtiene resultados levemente inferiores ($acc \approx 0.78$). NB posee mejor puntaje *precisión* (≈ 0.84), seguido de SVM y Logit con resultados similares (≈ 0.71 y ≈ 0.73 respectivamente) y ANN (≈ 0.65), lo cual quiere decir que NB posee el mejor desempeño en etiquetar casos positivos correctamente, sin embargo, también se observa que este posee el peor puntaje *recall* (≈ 0.341), lo cual implica que es inferior en la cantidad de etiquetas positivas que asigna con respecto al total de casos positivos de mortalidad.

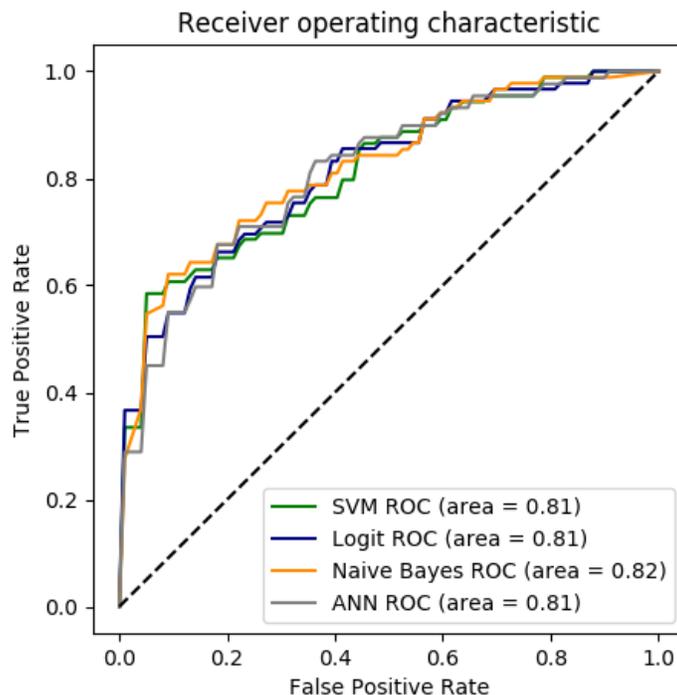
Adicionalmente, se realiza un análisis gráfico del desempeño del modelo mediante la figura 5.14, la cual representa la curva ROC promedio para cada algoritmo de clasificación de mortalidad de 72 horas. Por otro lado, la figura 5.15 corresponde a la comparación de la curva ROC promedio de los cuatro algoritmos utilizados en el análisis.

Figura 5.14: Curva ROC para cuatro clasificadores en modelo 72 horas. La curva azul opaca corresponde al promedio de cada curva generada por las iteraciones de la validación cruzada, las cuales se muestran en azul claro



Fuente: Elaboración Propia

Figura 5.15: Comparación curva ROC promedio para cuatro clasificadores en modelo 72 horas

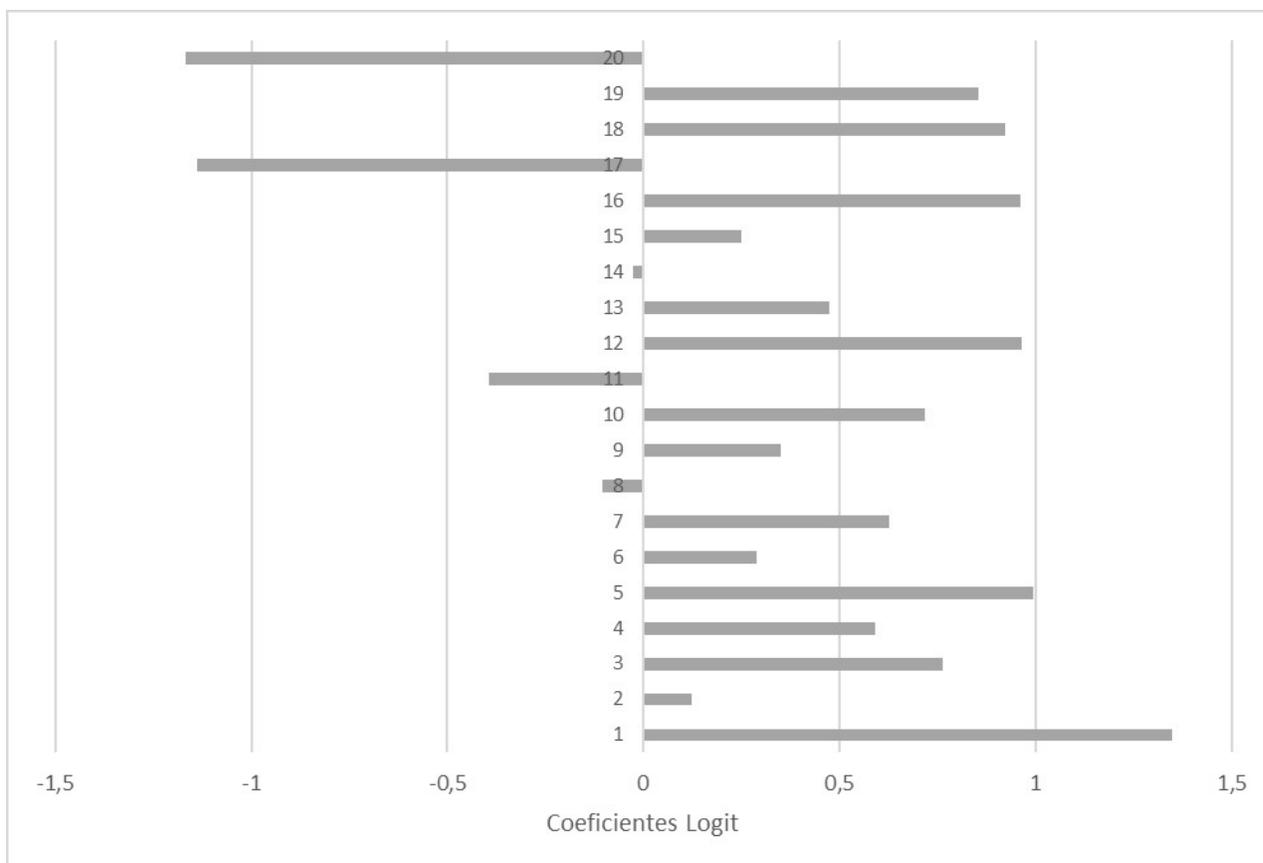


Fuente: Elaboración Propia

5.4.3. Interpretación Modelo 72 horas

Como se menciona en el capítulo 5.4.1, después de realizar el análisis para encontrar el umbral óptimo de selección de variables, el clasificador Logit se entrena con las 47 covariables más relevantes según el ranking de puntaje F ($p\text{-valor} \leq 0.05$), las cuales se detallan en el Apéndice A. La figura 5.16 muestra los coeficientes Logit correspondientes a las 20 variables más relevantes según puntaje F (el detalle con el nombre y descripción de variables se encuentra en la tabla 5.8). El análisis de estos coeficientes entrega una idea de cuáles son las variables más incidentes para determinar el desenlace del paciente una vez transcurridas las primeras 72 horas de hospitalización en UCI. Los coeficientes Logit para las 47 variables incluidas en el modelo se muestran en el Apéndice B.

Figura 5.16: Coeficientes Logit modelo 72 horas



Fuente: Elaboración Propia

De las variables seleccionadas, se destaca la alta influencia las siguientes variables protectoras ($\beta < 0$):

- **{20} Ácido Láctico bajo (máximo)** ($\beta \approx -1,17$): Indica un nivel de ácido láctico en la sangre menor al rango normal, para el máximo valor registrado durante las primeras 72 horas de hospitalización en UCI ($< 0,7 \text{ mmol/L}$).
- **{17} Ácido Úrico bajo (máximo)** ($\beta \approx -1,14$): Indica un nivel de ácido úrico en la sangre menor al rango normal, para el máximo valor registrado durante las primeras 72 horas de hospitalización en UCI ($< 2,5 \text{ F}, < 3,5 \text{ M}$).
- **{11} Hemoglobina baja (mínimo)** ($\beta \approx -0,39$): Indica bajos niveles de hemoglobina en la sangre ($< 7 \text{ gr/dL}$), para el mínimo valor registrado durante las primeras 72 horas de hospitalización en UCI, utilizado comúnmente como marcador de anemia. Niveles de hemoglobina bajos en la sangre está identificado en la literatura como factor de riesgo en pacientes críticos, por lo cual los resultados aquí obtenidos se califican como poco concluyentes.

- **{8} Bilirrubina alta (último)** ($\beta \approx -0,10$): Indica niveles de bilirrubina total muy altos ($> 2\text{mg/dl}$), en el último valor registrado durante las primeras 72 horas de hospitalización en UCI, lo cual es marcador de presencia de enfermedades hepáticas. Como se describe en la sección 5.2.3, este se ha definido como factor de riesgo y peores desenlaces en pacientes críticos, lo cual, sumado a la contradicción con respecto a los resultados obtenidos para el modelo de 48 horas ($\beta > 0$), este se considera poco concluyente.
- **{14} Saturación Oxígeno bajo (último)** ($\beta \approx 0,03$): Indica saturación de oxígeno en la sangre muy baja ($< 60\%$), en el mínimo valor registrado durante las primeras 72 horas de hospitalización en UCI. Como se describe en la sección 5.1.3, los niveles de oxígeno en la sangre han sido ampliamente utilizados como marcador de riesgo y predictor de mortalidad en pacientes críticos, lo cual, sumado a los hallazgos encontrados para el modelo de 24 horas ($\beta > 0$), este se considera poco concluyente

Se destaca la alta influencia como factor de riesgo de mortalidad de las siguientes variables ($\beta > 0,75$):

- **{1} Coagulopatía** ($\beta \approx 1,35$): Indica la presencia de alguna coagulopatía, como coagulación intravascular diseminada (CID) u otros defectos coagulatorios.
- **{5}/{10} Ácido Láctico alto (último)** ($\beta \approx 0,99/\beta \approx 0,98$): Indica la presencia de ácido láctico alto ($2,1 < AL \leq 4$) y muy alto ($AL > 4$) respectivamente, en la última medición registrada durante las primeras 72 horas de hospitalización. (para más detalles ver descripción en sección 5.1.3)
- **{12} Enfermedad Vascular Periférica** ($\beta \approx 0,97$): Indica la presencia de alguna enfermedad vascular periférica como diagnóstico secundario del paciente, por ejemplo, estenosis arterial.
- **{16} Ph alto (máximo)** ($\beta \approx 0,96$): Indica un nivel de Ph en la sangre muy alto ($Ph > 7,5$) en el máximo registro de las primeras 72 horas de hospitalización en UCI, afección conocida como alcalosis metabólica. El nivel de Ph anormal en la sangre, tanto si está por sobre (alcalosis) o bajo (acidosis) los rangos normales, es reconocido como un factor de gravedad en pacientes críticos, incluido por APACHE II en el cálculo del puntaje de riesgo[41].
- **{18} Potasio alto (promedio)** ($\beta \approx 0,92$): Indica un nivel de potasio por sobre el rango normal ($K > 5,1$) en el promedio de todas las mediciones disponible durante 72 horas de hospitalización en UCI (para más detalles ver descripción en sección 5.2.3).
- **{19} Carboxihemoglobina alta (último)** ($\beta \approx 0,85$): Indica un nivel de carboxihemoglobina en la sangre por sobre el rango normal ($> 2,5\%$). Al respecto no existe mucha documentación acerca de los efectos individuales que tiene esta variable sobre el desenlace clínico, pero el estudio de [55] concluye que tanto niveles muy bajos como niveles muy

altos de carboxihemoglobina se asocian con un aumento de la mortalidad en cuidados intensivos.

- **{3} Ph alto (promedio)** ($\beta \approx 0,76$): Indica un nivel de Ph en la sangre muy bajo ($Ph < 7,25$) en el promedio de todos los registros disponibles durante las primeras 72 horas de hospitalización en UCI, afección conocida como acidosis metabólica (para más detalles ver descripción en sección 5.1.3).

Se destaca una influencia moderada como factor de riesgo de mortalidad ($0,25 < \beta < 0,75$) de las siguientes variables:

- **{7} Enfermedad Hepática** ($\beta \approx 0,63$): Indica la presencia de alguna enfermedad hepática como diagnóstico secundario del paciente, como por ejemplo hepatitis crónica, fallo hepático, cirrosis, entre otras.
- **{4} LDH alto (máximo)** ($\beta \approx 0,59$): Indica un nivel de deshidrogenasa láctica (LDH) mayor al rango normal (> 618) en el máximo valor registrado durante las primeras 72 horas de hospitalización en UCI.
- **{13} Cáncer Estómago** ($\beta \approx 0,47$): Cáncer de estómago como diagnóstico principal.
- **{9} Días previos hospital (log)** ($\beta \approx 0,35$): Corresponde al logaritmo de la cantidad de días previos que el paciente ha estado hospitalizado antes de ingresar a UCI. Estos resultados sugieren que estadías previas más largas, es decir que el paciente tarda más tiempo en llegar al servicio, se asocia a peores desenlaces.
- **{6} Transaminasas altas (máximo)** ($\beta \approx 0,28$): Indica un nivel de alanino transaminasas muy alto ($AST > 126/206$ U/L (F/M)) en el máximo valor registrado durante las primeras 72 horas de hospitalización (para más detalles ver descripción en sección 5.1.3).
- **{15} Creatinina alta (promedio)** ($\beta \approx 0,25$): Indica niveles de creatinina por sobre el rango normal ($> 2\text{mg/dl}$), en el último valor registrado durante las primeras 72 horas de hospitalización en UCI. (para más detalles ver descripción en sección 5.1.3).

Se destaca una influencia menor de los siguientes factores de riesgo de mortalidad intrahospitalaria ($\beta < 0,25$)

- **{2} Plaquetas bajas (promedio)** ($\beta \approx 0,12$): Indica niveles de conteo de plaquetas por debajo del rango normal (< 150.000 K/UI), en el mínimo valor registrado durante las primeras horas de hospitalización en UCI (para más detalles ver descripción en sección 5.1.3).

5.5. Discusión

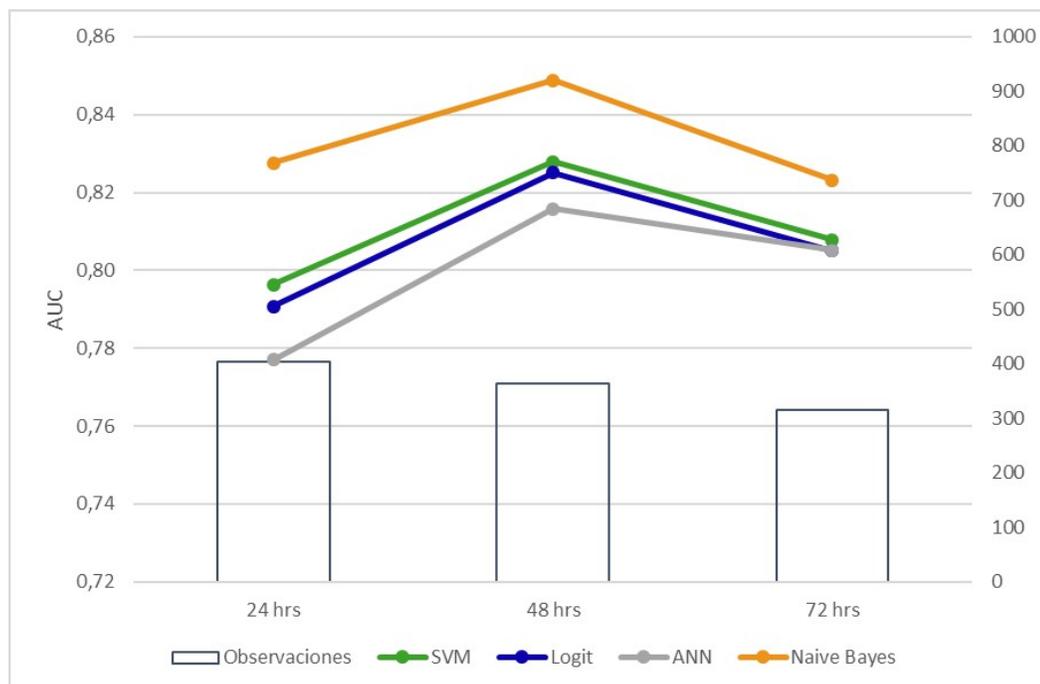
5.5.1. Análisis y Comparación de Resultados

La figura 5.17 muestra un gráfico que representa por un lado el promedio AUC de cada clasificador (líneas), y por otro lado número de observaciones disponibles (barras), para cada una de las configuraciones de predicción. Del gráfico se observa que, en el caso de todos los clasificadores, el promedio AUC de la validación cruzada es superior para el modelo de 48 horas. En segundo lugar, está el modelo 72 horas para SVM, Logit y ANN, y el modelo de 24 horas para NB; la diferencia en el desempeño de los distintos clasificadores disminuye al aumentar las horas de predicción. Respecto de lo anterior se concluye dos cosas:

En primer lugar, se concluye que el mejor desempeño del modelo de 48 horas probablemente se debe a que, a pesar de disminuir en un 10% la cantidad de observaciones disponibles para entrenar el modelo, se tiene mayor cantidad de datos disponibles para cada paciente, y, por lo tanto, un set de información más robusto.

Por otro lado, la disminución de desempeño al aumentar la predicción de 48 a 72 horas, da una pista de que, a pesar de contar con más información de cada paciente, la disminución en la cantidad de observaciones disponibles para entrenar el modelo (22% con respecto al de la base original), la tarea de predicción se hace más difícil. Al respecto se puede pensar que, en trabajos futuros, trabajar con bases de datos con un mayor número de observaciones podría aumentar bastante el desempeño de este tipo de modelos.

Figura 5.17: Comparación AUC



Fuente: Elaboración Propia

Con respecto a la interpretación de los resultados obtenidos después de entrenar el modelo Logit para cada una de las configuraciones de predicción, se destaca la influencia de los siguientes factores de riesgo, comunes para los tres modelos predictivos (recordando que en este análisis solo entran las 20 variables más relevantes para cada modelo):

- **Niveles de lactato elevados**, especialmente para aquellos pacientes que registran valores por sobre los 4 mmol/L.
- **Presencia de alguna coagulopatía** como diagnóstico secundario del paciente. Entre los diagnósticos más frecuentes para la cohorte estudiada destacan; coagulación intravascular diseminada, defectos de la coagulación no especificados, y trombocitopenia, afecciones que se manifiestan en la disminución de plaquetas en la sangre.
- En línea con el punto anterior, para los tres modelos, un **conteo muy bajo de plaquetas** (<80000 U/L) representa un factor asociado a mayor riesgo de mortalidad.
- **Presencia de alguna enfermedad del hígado** como diagnóstico secundario del paciente. Entre los diagnósticos más frecuentes para la cohorte estudiada destacan; insuficiencia hepática aguda, hipertensión portal, y enfermedad de hígado no especificada.
- **La cantidad de días de hospitalización previos a la admisión en UCI** (LOG_HOSP), según la cual estadías previas más largas se traducen en peores resultados una vez admitidos en UCI.

Por otro lado, se encontraron resultados contradictorios para las siguientes variables:

- **Aclaramiento de creatinina muy bajos** (<30 ml/min), es un factor de riesgo para el modelo de 24 horas, y un factor protector para el modelo de 48 horas. Como se menciona en la sección 5.1.3, este es un marcador utilizado para evaluar el funcionamiento de los riñones y para diagnosticar fallo renal agudo, siendo descrito por la literatura como un factor de predicción independiente de mortalidad. En trabajos futuros se sugiere evaluar los valores críticos y rangos de anormalidad.
- **Niveles de saturación de oxígeno muy bajos** (sO₂<60%) es un factor de riesgo para el modelo de 24 horas, y factor protector en modelo de 72 horas.
- **Niveles de bilirrubina elevados** (>2 mg/dl) es un factor de riesgo para el modelo 48 horas, y un factor protector en modelo 72 horas.

Con respecto a estos resultados contradictorios y/o poco concluyentes, se observa que en todos los casos para el modelo de período más corto los resultados son más concordantes con la literatura revisada, de lo cual se concluye que, al disminuir la cantidad de observaciones disponibles, se debilita la relación estadística entre cada predictor y la variable objetivo, lo cual se puede observar en la disminución de puntajes F a medida que aumenta el período de descripción.

Por otro lado, se debe tener en cuenta que se utilizan distintos número de variables para los tres modelos propuestos para entrenar el modelo Logit (128, 114 y 47). Esto puede incidir en que los pesos asignados a cada variable pueden variar con la presencia/ausencia de otras variables. Esto se ve claramente reflejado en el hecho de que los resultados contradictorios se dan por lo general en variables con coeficientes pequeños ($abs(\beta) \leq 0.25$).

Con respecto a los resultados poco concluyentes, en este caso los que contradicen lo que

demuestra la literatura revisada o entre los distintos modelos propuestos, para trabajos futuros se sugiere:

1. Comparar resultados utilizando las mismas variables en cada modelo. Esto además tiene la ventaja de poder observar la evolución de los distintos factores de riesgo.
2. Incluir en el análisis el nivel de significancia de cada variable en el modelo, a través de, por ejemplo, el puntaje t y correspondiente p-valor. Esto no se realiza en el presente trabajo ya que el paquete estadístico que se utiliza (Python Scikit-Learn), no incluye este tipo de herramientas.
3. Trabajar con bases de datos con mayor número de observaciones para obtener resultados más robustos y concluyentes.
4. Evaluar el uso de distintos valores críticos en rangos de normalidad para valores de laboratorio.

5.5.2. Opinión Experta

Además de la comparación e interpretación de resultados, estos se validan con la contraparte médica del HCUCH para corroborar sobre el valor y sentido clínico de los resultados y los factores de riesgo que se desprenden de estos. La discusión se resume en los siguientes puntos.

- En términos generales, se destaca que el desempeño predictivo de los modelos presentados en este trabajo (AUC 0.78 – 0.85) es similar al desempeño que otros estudios internacionales reportan sobre la validación de los sistemas de scoring clásicos más utilizados en el HCUCH y en general en el sistema de salud chileno: APACHE II y SOFA, con la diferencia de que los sistemas nombrados se emplean en general para el análisis global de las UCI, más que para asignar niveles de riesgo para cada paciente individualmente, lo cual es una ventaja para el modelo propuesto. Se destacan además otras ventajas y aportes de este trabajo:
 - Es un modelo potencialmente útil para crear sistemas predictivos automáticos y dinámicos en el tiempo. Uno de los limitantes de APACHE II y la razón por la cual no se usa tanto en la práctica, es que el cálculo toma tiempo ya que generalmente no se tienen todas las variables a la mano. Por otro lado, el sistema es estático, siendo diseñado solo para las primeras 24 horas transcurridas después de la admisión. Crear sistemas dinámicos facilitaría la caracterización de la evolución del riesgo del paciente, además de otorgar flexibilidad en el momento del cálculo, sin restringirse a ciertas ventanas temporales.
 - El modelo representa un entendimiento de una gran cantidad de variables en un solo indicador de riesgo. Los médicos se enfrentan a una gran y diversa cantidad de información por cada paciente tratado, dispersa en distintos sistemas y bajo distintas formas, lo cual resulta difícil de manejar y ponderar.
 - Se considera un análisis metodológico robusto, muy por sobre de los empleados habitualmente en el diseño de sistemas pronósticos en UCI.

- Ventaja de ser un modelo construido a partir de población específica de pacientes chilenos, lo cual no se había realizado antes.
- Ventaja de ser un modelo construido a partir de población específica de pacientes adultos mayores, donde existen menos estudios específicos.
- En relación a los factores de riesgo que se desprenden de la interpretación de los modelos, se concluye que son biológicamente plausibles, algunos de ellos ampliamente reconocidos como el lactato elevado, la presencia de coagulopatía, presencia de enfermedades hepáticas, plaquetas bajas, Ph bajo, creatinina alta.
Aparecen otras variables de riesgo que se mencionan o documentan con menor frecuencia pero que son razonables desde el punto de vista clínico, como la cantidad de días de hospital previos al ingreso a UCI.
Por otro lado, aparecen variables de riesgo de mortalidad menos documentadas y que por lo tanto son de menos interés del médico en el tratamiento individual de cada paciente, hallazgos que en trabajos futuros sería útil seguir investigando. Ejemplo de estas son algunos resultados de laboratorio, como las transaminasas (AST), LDH, potasio, albúmina, procalcitonina, entre otros.
- Se destaca sobre la novedad y potencial utilidad de incluir indicadores de ausencia que demostraron ser factores de riesgo o protección, por ejemplo, la ausencia de examen de procalcitonina. Este análisis podría extenderse a otros indicadores de procedimientos y tratamientos.
- Se destaca la limitante que representa la falta de parámetros clínicos habituales de las primeras 48 horas de estadía en UCI (temperatura, presión arterial, frecuencia cardiaca, etc), por la ausencia de estos en ficha clínica, las cuales sería importante incluir en trabajos futuros.

Capítulo 6

Conclusiones y Trabajo Futuro

6.1. Conclusiones

Esta investigación propone la construcción de un modelo para predecir mortalidad en pacientes críticos adultos mayores admitidos en UCI de un hospital académico chileno, a través de información específica de cada paciente, almacenada bajo la forma de bases de datos disponibles en los sistemas de información del hospital.

Los objetivos se cumplieron llevando a cabo un análisis metodológico profundo que contempló la combinación de varias técnicas de análisis, procesamiento y modelamiento de datos para obtener los resultados deseados, producto de lo cual se desarrolló un modelo de predicción que entrega un indicador de riesgo para cada paciente individual, para tres intervalos de tiempo de hospitalización: las primeras 24, 48 y 72 horas de hospitalización en UCI, demostrando la obtención de resultados satisfactorios y comparables a los resultados documentados sobre el uso de las herramientas de scoring que se utilizan hoy en día, teniendo además la ventaja de que, con un mayor desarrollo, afinamiento, y entrenamiento de este tipo de modelos, en un largo plazo se podrían llegar a desarrollar herramientas de apoyo a la decisión en tiempo real, basado en procesos de aprendizaje que utilicen la data almacenada en registro electrónico de manera automática. Por otro lado, se destaca acerca del carácter novedoso de la creación de un modelo predictivo de riesgo a partir de población chilena.

De lo anterior se desprende el valor y aporte de esta investigación, que radica en la demostración y cumplimiento de la hipótesis de esta, la cual propone que es posible predecir desenlaces clínicos a partir de la data almacenada en registros electrónicos de salud para una cohorte de pacientes chilenos, dando aliento a la generación de nuevas investigaciones que sigan en línea con los resultados aquí expuestos, de forma de crear y potenciar suficiente evidencia para despertar el interés tanto de los médicos, investigadores, hospitales y autoridades de la salud pública, con el objetivo de hacer del análisis y almacenamiento de datos una prioridad en temas de salud pública.

En este sentido, es importante destacar que la investigación tiene la ventaja de estar enmarcado en un centro de investigación académico (WIC), que posee tanto los recursos y el capital

humano para impulsar una vertical de trabajo a partir de los hallazgos de este trabajo, a través de futuras investigaciones y trabajos de título, capitalizando así el conocimiento generando una continuidad y aprendizaje colaborativo.

Con respecto a la metodología, “Knowledge Discovery in DataBases” (KDD), se desprende que aportó en dar un marco de trabajo que otorgó una linealidad y orden claros al trabajo y a los pasos necesarios para obtener los resultados que se esperaban. Por otro lado, en cada uno de los pasos, el procesamiento y análisis que se realizó sobre los datos fue producto de una investigación individual y de la orientación en conocimiento experto aportado por la contraparte médica, en este caso el doctor Eduardo Tobar. De esto se concluye que, por un lado, el resultado de este trabajo es producto de un proceso de aprendizaje único que puede derivar en una gama amplia de variantes a lo aquí propuesto, lo que significa una oportunidad de seguir iterando sobre la metodología y potencialmente mejorando el desempeño del modelo.

Con respecto a los resultados del modelamiento, se demostró que los modelos construidos obtuvieron buen desempeño, demostrado además mediante el análisis de sensibilidad que implicó la calibración del modelamiento utilizando distintas configuraciones de predicción, con promedios AUC bordeando 0.78-0.83, 0.82-0.85 y 0.81-0.82 para los modelos de 24, 48 y 72 horas respectivamente, observando diferencias con respecto a cada clasificador utilizado y entre las distintas configuraciones de predicción, obteniendo mejores desempeños con el clasificador Naïve Bayes para los tres casos, y mejores desempeños para el modelo de 48 horas, para todos los clasificadores. Otra de las ventajas de realizar este análisis de sensibilidad es que otorga la oportunidad de crear predicciones dinámicas en el tiempo, pudiendo seguir el estado de gravedad del paciente a medida que se acumula la data. La única desventaja encontrada con respecto a esto es que, como no todos los pacientes poseen estadías largas, a medida que se alarga el período de predicción, disminuye la cantidad de observaciones, y, por lo tanto, el poder predictivo del modelo.

En cuanto a la interpretación de los resultados obtenidos producto del entrenamiento del clasificador Logit, en términos generales se observan patrones clínicamente plausibles e intuitivos, encontrando algunos factores que sería interesante seguir investigando en futuros trabajos, como, por ejemplo, la ausencia de eventos y órdenes clínicas, variables derivadas como la cantidad de días de hospitalización previas a UCI, algunos resultados de laboratorios menos mencionados en la literatura, entre otros. Con respecto a los resultados contradictorios, se concluye sobre el beneficio que significaría trabajar con bases de datos con mayor número de observaciones, y de realizar un análisis de significancia en la interpretación del modelo. Además, se pone en duda sobre la viabilidad de comparar modelos que usan sets de variables distintos, lo cual puede ser una de las razones de los resultados contradictorios.

Por otro lado, se comprende acerca de la complejidad de llevar a cabo tareas de modelamiento predictivo y de interpretación estadística en el mismo marco de trabajo. Esto se refleja en que, por un lado, tanto el lenguaje como los paquetes computacionales utilizados (Python) están más enfocados en el modelamiento que en el análisis estadístico, permitiendo reportar sobre la influencia de algunas de las variables como factores de riesgo, pero no pudiendo detallar sobre la significancia estadística de las variables que se analizaron con respecto a la variable objetivo. Por otro lado, la complejidad de esto radica en la cantidad de variables incluidas en los modelos 128, 114 y 47 para los modelos de 24, 48 y 72, haciendo más difícil la interpretación

sobre el rol individual que tiene cada variable en la predicción de riesgo.

Por último, se concluye sobre la complejidad y el desafío que significa llevar a cabo el proceso de extracción de conocimiento a partir de datos, aplicado en disciplinas tan ajenas a la ingeniería como lo es la medicina, tendiendo que llevarse a cabo un profundo proceso de aprendizaje acerca de los procesos y causas que generaron los datos con los que se está trabajando, para poder plantear interpretaciones y nuevos conocimientos potencialmente útiles. Por lo anterior, se destaca la importancia de la formación de equipos multidisciplinarios en el desarrollo de este tipo de herramientas.

6.2. Trabajo Futuro

1. Construir bases de datos con mayor número de observaciones. Como se menciona en capítulos anteriores, esto podría mejorar el desempeño y ofrecer resultados más robustos y concluyentes
2. Construir una base de datos multi-céntrica. Un sesgo importante del presente estudio es que incluye información sobre pacientes admitidos en un solo hospital.
3. En el procesamiento de los datos, evaluar distintos métodos de clasificación y representación de la data. Por ejemplo:
 - Para los datos de laboratorio se podría evaluar utilizar variables continuas mediante la imputación de datos faltantes, o realizar un análisis de sensibilidad utilizando distintos rangos de normalidad.
 - Para los medicamentos, evaluar otros métodos de clasificación, y/o crear variables que informen sobre la interacción entre medicamentos.
4. En el modelamiento realizar un análisis de sensibilidad evaluando el entrenamiento de distintos hiperparámetros característicos de los clasificadores. Por temas de extensión y alcance se dejó fuera de este trabajo, pero llevar a cabo un análisis de este tipo podría mejorar los resultados aquí mostrados.
5. Incluir variables fisiológicas básicas descritas como factores de riesgo de mortalidad: T, ritmo cardiaco, ritmo respiratorio y presión arterial, las cuales no se entran en este análisis por no estar disponibles de forma electrónica.
6. De estar disponible, incluir Escala de Coma de Glasgow (GCS) como variable en el modelo, ampliamente descrito como factor predictor de mortalidad. En este análisis no se incluye por no estar disponible, sin embargo, se podría considerar en estudios prospectivos.
7. En el caso de desarrollar modelos predictivos de riesgo en adultos mayores, incluir factores específicos de este grupo etario, lo cual se reconoce como una debilidad para este estudio, en el cual se incluyen las comorbilidades presenes, pero no se considera el estatus funcional y cognitivo, descrito en el capítulo 1 como factores de riesgo importantes que podrían considerarse en estudios prospectivos.
8. Desarrollar modelos específicos para cada diagnóstico, por ejemplo, algunos de los más comunes en pacientes críticos AM. Para esto se necesitarían bases de datos más robustas en observaciones.
9. Para confirmar la superioridad del modelo con respecto a los sistemas de scoring existentes, se podría comparar el desempeño con aquel ofrecido por APACHE II para la misma

cohorte estudiada. No fue posible en este trabajo por la falta de medidas fisiológicas nombradas en el punto 5.

10. Utilizar técnicas de text mining. Otra de las debilidades de este trabajo es que los diagnósticos clínicos que se atribuyen a cada paciente generalmente son procesados después de que el paciente fue dado de alta, y los códigos CIE-10 son asignados según las descripciones de las notas médicas de doctores y enfermeras disponibles en ficha clínica. Dado esto, se han desarrollado algunos estudios de este tipo que utilizan text-mining para interpretar las notas médicas y traducirlas en diagnósticos clínicos.
11. De realizarse un análisis estadístico más profundo, en trabajos futuros se sugiere por un lado crear modelos con un número acotado de variables, y utilizar técnicas estadísticas de evaluación de significancia.

Capítulo 7

Bibliografía

- [1] RIFLE criteria for acute kidney injury are associated with hospital mortality in critically ill patients: a cohort analysis. 10:R73, May 2006.
- [2] Paul D. Allison. *Missing Data*. SAGE Publications, 2002. Google-Books-ID: ZtYArHXjpB8C.
- [3] Ruben Amarasingham, Anne-Marie J. Audet, David W. Bates, I. Glenn Cohen, Martin Entwistle, G. J. Escobar, Vincent Liu, Lynn Etheredge, Bernard Lo, Lucila Ohno-Machado, Sudha Ram, Suchi Saria, Lisa M. Schilling, Anand Shahi, Walter F. Stewart, Ewout W. Steyerberg, and Bin Xie. Consensus Statement on Electronic Health Predictive Analytics: A Guiding Framework to Address Challenges. *eGEMs*, 4(1), March 2016.
- [4] Ruben Amarasingham, Rachel E. Patzer, Marco Huesch, Nam Q. Nguyen, and Bin Xie. Implementing electronic health care predictive analytics: considerations and challenges. *Health Affairs (Project Hope)*, 33(7):1148–1154, July 2014.
- [5] Finn H. Andersen, Hans Flaatten, Pål Klepstad, Ulla Romild, and Reidar Kvåle. Long-term survival and quality of life after intensive care for patients 80 years of age or older. *Annals of Intensive Care*, 5(1):53, December 2015.
- [6] Ian M. Ball, Sean M. Bagshaw, Karen E. A. Burns, Deborah J. Cook, Andrew G. Day, Peter M. Dodek, Demetrios J. Kutsogiannis, Sangeeta Mehta, John G. Muscedere, Henry T. Stelfox, Alexis F. Turgeon, George A. Wells, and Ian G. Stiell. A clinical prediction tool for hospital mortality in critically ill elderly patients. *Journal of Critical Care*, 35:206–212, October 2016.
- [7] Riccardo Bellazzi and Blaz Zupan. Predictive data mining in clinical medicine: current issues and guidelines. *International Journal of Medical Informatics*, 77(2):81–97, February 2008.
- [8] Richard Bellman and Richard Ernest Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, 1961. Google-Books-ID: POAmAAAAMAAJ.
- [9] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational*

learning theory, pages 144–152. ACM, 1992.

- [10] Nathan E. Brummel, Michele C. Balas, Alessandro Morandi, Lauren E. Ferrante, Thomas M. Gill, and E. Wesley Ely. Understanding and reducing disability in older adults following critical illness. *Critical Care Medicine*, 43(6):1265–1275, June 2015.
- [11] Mary E. Charlson, Peter Pompei, Kathy L. Ales, and C. Ronald MacKenzie. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases*, 40(5):373–383, January 1987.
- [12] Y. Chen and H. Yang. Heterogeneous postsurgical data analytics for predictive modeling of mortality risks in intensive care units. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4310–4314, August 2014.
- [13] Yi-Wei Chen and Chih-Jen Lin. Combining SVMs with Various Feature Selection Strategies. In Isabelle Guyon, Masoud Nikravesh, Steve Gunn, and Lotfi A. Zadeh, editors, *Feature Extraction*, number 207 in Studies in Fuzziness and Soft Computing, pages 315–324. Springer Berlin Heidelberg, 2006. DOI: 10.1007/978-3-540-35488-8_13.
- [14] Departamento de Estadísticas e Información de Salud (DEIS).
- [15] Ministerio de Salud Gobierno de Chile. Encuesta nacional de salud, 2010.
- [16] Superintendencia de Salud Gobierno de Chile. Impacto del envejecimiento en el gasto en salud: Chile 2002-2020, 2006.
- [17] Moritoki Egi, Rinaldo Bellomo, Edward Stachowski, Craig J. French, Graeme K. Hart, Gopal Taori, Colin Hegarty, and Michael Bailey. Hypoglycemia and Outcome in Critically Ill Patients. *Mayo Clinic Proceedings*, 85(3):217–224, March 2010.
- [18] Anne Elixhauser, Claudia Steiner, D. Robert Harris, and Rosanna M. Coffey. Comorbidity Measures for Use with Administrative Data. *Medical Care*, 36(1):8–27, 1998.
- [19] Andrés Esteban, Antonio Anzueto, Fernando Frutos, Inmaculada Alía, Laurent Brochard, Thomas E. Stewart, Salvador Benito, Scott K. Epstein, Carlos Apezteguía, Peter Nightingale, Alejandro C. Arroliga, Martin J. Tobin, and for the Mechanical Ventilation International Study Group. Characteristics and Outcomes in Adult Patients Receiving Mechanical Ventilation: A 28-Day International Study. *JAMA*, 287(3):345–355, January 2002.
- [20] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3):37, March 1996.
- [21] Lauren E. Ferrante, Margaret A. Pisani, Terrence E. Murphy, Evelyne A. Gahbauer, Linda S. Leo-Summers, and Thomas M. Gill. Functional trajectories among older persons before and after critical illness. *JAMA internal medicine*, 175(4):523–529, April 2015.
- [22] Simon J. Finney, Cornelia Zekveld, Andi Elia, and Timothy W. Evans. Glucose Control and Mortality in Critically Ill Patients. *JAMA*, 290(15):2041–2047, October 2003.

- [23] Gersende Fort and Sophie Lambert-Lacroix. Classification using partial least squares with penalized logistic regression. *Bioinformatics*, 21(7):1104–1111, April 2005.
- [24] Jean-Roger Le Gall, Stanley Lemeshow, and Fabienne Saulnier. A New Simplified Acute Physiology Score (SAPS II) Based on a European/North American Multicenter Study. *JAMA*, 270(24):2957–2963, December 1993.
- [25] Francisca García Lizana, José Luis Manzano Alonso, and Pedro Saavedra Santana. Mortalidad y calidad de vida al año del alta de una unidad de cuidados intensivos en pacientes mayores de 65 años. *Medicina Clínica*, pages 521–525.
- [26] García Ubilla, Arnol. Análisis de datos y búsqueda de patrones en aplicaciones médicas. *Repositorio Académico - Universidad de Chile*, 2015.
- [27] Marzyeh Ghassemi, Tristan Naumann, Finale Doshi-Velez, Nicole Brimmer, Rohit Joshi, Anna Rumshisky, and Peter Szolovits. Unfolding Physiological State: Mortality Modelling in Intensive Care Units. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14*, pages 75–84, New York, NY, USA, 2014. ACM.
- [28] Marzyeh Ghassemi, Marco A.F. Pimentel, Tristan Naumann, Thomas Brennan, David A. Clifton, Peter Szolovits, and Mengling Feng. A Multivariate Timeseries Modeling Approach to Severity of Illness Assessment and Forecasting in ICU with Sparse, Heterogeneous Clinical Data. *Proceedings of the ... AAAI Conference on Artificial Intelligence. AAAI Conference on Artificial Intelligence*, 2015:446–453, January 2015.
- [29] S. Gupta, S. Rana, D. Phung, and S. Venkatesh. Keeping up with Innovation: A Predictive Framework for Modeling Healthcare Data with Evolving Clinical Interventions. In *Proceedings of the 2014 SIAM International Conference on Data Mining, Proceedings*, pages 235–243. Society for Industrial and Applied Mathematics, April 2014. DOI: 10.1137/1.9781611973440.27 DOI: 10.1137/1.9781611973440.27.
- [30] Dominik G. Haider, Gregor Lindner, Michael Wolzt, Sufian S. Ahmad, Thomas Sauter, Alexander Benedikt Leichtle, Georg-Martin Fiedler, Valentin Fuhrmann, and Aristomenis K. Exadaktylos. Hyperphosphatemia Is an Independent Risk Factor for Mortality in Critically Ill Patients: Results from a Cross-Sectional Study. *PLoS ONE*, 10(8), August 2015.
- [31] PubMed Health. Drug names and classes, 2015.
- [32] Jeroen Hermanides, Robert J. Bosman, Titia M. Vriesendorp, Ron Dotsch, Frits R. Rosendaal, Durk F. Zandstra, Joost B. L. Hoekstra, and J. Hans DeVries. Hypoglycemia is associated with intensive care unit mortality. *Critical Care Medicine*, 38(6):1430–1434, June 2010.
- [33] Lara Hessels, Miriam Hoekstra, Lisa J. Mijzen, Mathijs Vogelzang, Wim Dieperink, Annetie Oude Lansink, and Maarten W. Nijsten. The relationship between serum potassium, potassium variability and in-hospital mortality in critically ill patients and a before-after analysis on the impact of computer-assisted potassium control. *Critical Care (London, England)*, 19:4, January 2015.

- [34] Jaime Hidalgo A. ¿Por qué son necesarias unidades geriátricas de agudos en el sistema de salud chileno? *Rev. Hosp. Clin. Univ. Chile*, 26(4):306–313, 2015.
- [35] M. Hoogendoorn, A. el Hassouni, K. Mok, M. Ghassemi, and P. Szolovits. Prediction using patient comparison vs. modeling: A case study for mortality prediction. pages 2464–2467, August 2016.
- [36] Caleb W Hug and Peter Szolovits. ICU Acuity: Real-time Models versus Daily Models. *AMIA Annual Symposium Proceedings*, 2009:260–264, 2009.
- [37] The NICE-SUGAR Study Investigators. Hypoglycemia and Risk of Death in Critically Ill Patients. *New England Journal of Medicine*, 367(12):1108–1118, September 2012.
- [38] Jens Ulrik Jensen, Lars Heslet, Tom Hartvig Jensen, Kurt Espersen, Peter Steffensen, and Michael Tvede. Procalcitonin increase in early identification of critically ill patients at high risk of mortality. *Critical Care Medicine*, 34(10):2596–2602, October 2006.
- [39] Peter B. Jensen, Lars J. Jensen, and Søren Brunak. Mining electronic health records: towards better research applications and clinical care. *Nature Reviews Genetics*, 13(6):395–405, June 2012.
- [40] Joseph A. Johnston, Douglas P. Wagner, Stephen Timmons, Deborah Welsh, Joel Tsevat, and Marta L. Render. Impact of Different Measures of Comorbid Disease on Predicted Mortality of Intensive Care Unit Patients. *Medical Care*, 40(10):929–940, 2002.
- [41] W. A. Knaus, E. A. Draper, D. P. Wagner, and J. E. Zimmerman. APACHE II: a severity of disease classification system. *Critical Care Medicine*, 13(10):818–829, October 1985.
- [42] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Stanford, CA, 1995.
- [43] S. B. Kotsiantis. Supervised Machine Learning: A Review of Classification Techniques. In *Proceedings of the 2007 Conference on Emerging Artificial Intelligence Applications in Computer Engineering: Real World AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*, pages 3–24, Amsterdam, The Netherlands, The Netherlands, 2007. IOS Press.
- [44] James S. Krinsley, Marcus J. Schultz, Peter E. Spronk, Robin E. Harmsen, Floris van Braam Houckgeest, Johannes P. van der Sluijs, Christian Mélot, and Jean Charles Preiser. Mild hypoglycemia is independently associated with increased mortality in the critically ill. *Critical Care*, 15:R173, July 2011.
- [45] James Stephen Krinsley. Effect of an Intensive Glucose Management Protocol on the Mortality of Critically Ill Adult Patients. *Mayo Clinic Proceedings*, 79(8):992–1000, August 2004.
- [46] Harlan M. Krumholz. Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. *Health Affairs (Project Hope)*, 33(7):1163–1170, July 2014.

- [47] Ole Kruse, Niels Grunnet, and Charlotte Barfod. Blood lactate as a predictor for in-hospital mortality in patients admitted acutely to hospital: a systematic review. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, 19:74, December 2011.
- [48] Guillaume Leblanc, Ariane Boumendil, and Bertrand Guidet. Ten things to know about critically ill elderly patients. *Intensive Care Medicine*, 43(2):217–219, February 2017.
- [49] Li-wei Lehman, Mohammed Saeed, William Long, Joon Lee, and Roger Mark. Risk stratification of ICU patients using topic models inferred from unstructured progress notes. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2012:505–511, 2012.
- [50] Stanley Lemeshow, Daniel Teres, Janelle Klar, Jill Spitz Avrunin, Stephen H. Gehlbach, and John Rapoport. Mortality Probability Models (MPM II) Based on an International Cohort of Intensive Care Unit Patients. *JAMA*, 270(20):2478–2486, November 1993.
- [51] Jau-Huei Lin and Peter J. Haug. Exploiting missing clinical data in Bayesian network modeling for predicting medical problems. *Journal of Biomedical Informatics*, 41(1):1–14, February 2008.
- [52] Roderick J. A. Little and Donald B. Rubin. *Statistical Analysis with Missing Data*. John Wiley & Sons, August 2014. Google-Books-ID: AyVeBAAAQBAJ.
- [53] Yen-Fu Luo and Anna Rumshisky. Interpretable Topic Features for Post-ICU Mortality Prediction. *AMIA Annual Symposium Proceedings*, 2016:827–836, February 2017.
- [54] Matjaž Majnik and Zoran Bosnić. ROC analysis of classifiers in machine learning: A survey. *Intelligent Data Analysis*, 17(3):531–558, January 2013.
- [55] Daniel D. Melley, Simon J. Finney, Androula Elia, Anna L. Lagan, Gregory J. Quinlan, and Timothy W. Evans. Arterial carboxyhemoglobin level and outcome in critically ill patients. *Critical Care Medicine*, 35(8):1882–1887, August 2007.
- [56] Mariano E. Menendez, Valentin Neuhaus, C. Niek van Dijk, and David Ring. The Elixhauser comorbidity method outperforms the Charlson index in predicting inpatient death after orthopaedic surgery. *Clinical Orthopaedics and Related Research*, 472(9):2878–2886, September 2014.
- [57] Bruce A Molitoris. Measuring glomerular filtration rate in the intensive care unit: no substitutes please. *Critical Care*, 17(5):181, 2013.
- [58] Karel G. M. Moons, Douglas G. Altman, Johannes B. Reitsma, John P. A. Ioannidis, Petra Macaskill, Ewout W. Steyerberg, Andrew J. Vickers, David F. Ransohoff, and Gary S. Collins. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of Internal Medicine*, 162(1):W1–73, January 2015.
- [59] Benson Mwangi, Tian Siva Tian, and Jair C. Soares. A review of feature reduction techniques in neuroimaging. *Neuroinformatics*, 12(2):229–244, April 2014.

- [60] Nozomi Nori, Hisashi Kashima, Kazuto Yamashita, Hiroshi Ikai, and Yuichi Imanaka. Simultaneous Modeling of Multiple Diseases for Mortality Prediction in Acute Hospital Care. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, pages 855–864, New York, NY, USA, 2015. ACM.
- [61] F. Philippart, A. Vesin, C. Bruel, A. Kpodji, B. Durand-Gasselín, P. Garçon, M. Levy-Soussan, J. L. Jagot, N. Calvo-Verjat, J. F. Timsit, B. Misset, and M. Garrouste-Orgeas. The ETHICA study (part I): elderly's thoughts about intensive care unit admission for life-sustaining treatments. *Intensive Care Medicine*, 39(9):1565–1573, September 2013.
- [62] Therese D. Pigott. A Review of Methods for Missing Data. *Educational Research and Evaluation*, 7(4):353–383, December 2001.
- [63] Hude Quan, Vijaya Sundararajan, Patricia Halfon, Andrew Fong, Bernard Burnand, Jean-Christophe Luthi, L. Duncan Saunders, Cynthia A. Beck, Thomas E. Feasby, and William A. Ghali. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical Care*, 43(11):1130–1139, November 2005.
- [64] Yvan Saeys, Iñaki Inza, and Pedro Larrañaga. A review of feature selection techniques in bioinformatics. *Bioinformatics (Oxford, England)*, 23(19):2507–2517, October 2007.
- [65] Alvaro Silva, Paulo Cortez, Manuel Filipe Santos, Lopes Gomes, and José Neves. Mortality assessment in intensive care units via adverse events using artificial neural networks. *Artificial Intelligence in Medicine*, 36(3):223–234, March 2006.
- [66] George C. M. Siontis, Ioanna Tzoulaki, and John P. A. Ioannidis. Predicting Death: An Empirical Evaluation of Predictive Tools for Mortality. *Archives of Internal Medicine*, 171(19):1721–1726, October 2011.
- [67] Vinko Tomicic, Mauricio Espinoza, Max Andresen, Jorge Molina, Mario Calvo, Héctor Ugarte, Jorge Godoy, Sergio Gálvez, Juan Carlos Maurelia, Iris Delgado, and Andrés Esteban. Características de los pacientes que reciben ventilación mecánica en unidades de cuidados intensivos: primer estudio multicéntrico chileno. *Revista médica de Chile*, 136(8):959–967, August 2008.
- [68] Esther V. Uijtendaal, Jeannette E. F. Zwart-van Rijkom, Dylan W. de Lange, Arief Lalmohamed, Wouter W. van Solinge, and Toine C. G. Egberts. Influence of a strict glucose protocol on serum potassium and glucose concentrations and their association with mortality in intensive care patients. *Critical Care (London, England)*, 19:270, June 2015.
- [69] S. Vanderschueren, A. De Weerd, M. Malbrain, D. Vankersschaever, E. Frans, A. Wilmer, and H. Bobbaers. Thrombocytopenia and prognosis in intensive care. *Critical Care Medicine*, 28(6):1871–1876, June 2000.
- [70] J.-L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, C. K. Reinhart, P. M. Suter, and L. G. Thijs. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Medicine*, 22(7):707–710, July 1996.
- [71] Jean-Louis Vincent, Marc-Jacques Dubois, Roberta J. Navickis, and Mahlon M. Wilkes. Hy-

poalbuminemia in Acute Illness: Is There a Rationale for Intervention? *Annals of Surgery*, 237(3):319–334, March 2003.

- [72] Monica Linea Vold, Ulf Aasebø, Tom Wilsgaard, and Hasse Melbye. Low oxygen saturation and mortality in an adult cohort: the Tromsø study. *BMC pulmonary medicine*, 15:9, February 2015.
- [73] U. Waheed, P. Williams, S. Brett, G. Baldock, and N. Soni. White cell count and intensive care unit outcome. *Anaesthesia*, 58(2):180–182, February 2003.
- [74] Brian J. Wells, Kevin M. Chagin, Amy S. Nowacki, and Michael W. Kattan. Strategies for Handling Missing Data in Electronic Health Record Derived Data. *eGEMs*, 1(3), December 2013.
- [75] Jionglin Wu, Jason Roy, and Walter F. Stewart. Prediction Modeling Using EHR Data: Challenges, Strategies, and a Comparison of Machine Learning Approaches. *Medical Care*, 48(6):S106–S113, 2010.
- [76] S Zampieri, P Bettonte, M Ortolani, G Frison, V Schweiger, L Gottin, and E Polati. Procalcitonin as prognostic marker of mortality. *Critical Care*, 17(Suppl 2):P29, 2013.
- [77] G. Peter Zhang. Neural Networks For Data Mining. In *Data Mining and Knowledge Discovery Handbook*, pages 419–444. Springer, Boston, MA, 2009. DOI: 10.1007/978-0-387-09823-4_21.

Apéndice A

Resultados Análisis Puntaje F

En esta sección del apéndice se adjuntan las variables más relevantes para el modelo de 24 horas, junto a su correspondiente puntaje F.

A.1. Variables y puntaje F para corte p-valor=0.3 para modelo 24 horas

Tabla A.1: Puntaje F variables relevantes, modelo 24 horas

Variable	Puntaje F	p-valor
LAST_AL_HIGH2	68,1283	2,236E-10
EX21	32,1673	2,711E-07
LAST_PH_LOW1	32,1673	2,711E-07
EX14	23,1544	2,120E-06
MEAN_K_HIGH	23,1011	2,177E-06
MEAN_AST_HIGH3	22,2597	3,294E-06
LAST_PLT_LOW1	20,4564	8,046E-06
ING_LDH_HIGH	19,5085	1,290E-05
MAX_NA_HIGH	17,6185	3,33E-05
VM	16,113	0,0001
MIN_ALB_LOW1	12,5539	0,0004
profilaxisgas	12,1518	0,0005
LAST_P_HIGH2	12,1518	0,0005
LOG_HOSP	12,0819	0,0006
LAST_BUN_HIGH2	11,9962	0,0006
ING_GLC_LOW	11,5822	0,0007
MIN_WBC_LOW1	11,2141	0,0009
MIN_CRC_LOW1	10,5313	0,0013
LAST_CR_HIGH2	9,7587	0,0019
MIN_SO2_LOW1	9,5799	0,0021

Variable	Puntaje F	p-valor
LAST_UA_HIGH	8,9429	0,003
LAST_WBC_HIGH1	8,5723	0,0036
EX5	8,0498	0,0048
MAX_CRC_LOW2	7,9173	0,0051
MIN_CR_HIGH1	7,5589	0,0062
C16	7,0657	0,0082
ING_PCT_HIGH3	7,0123	0,0084
MEAN_ALP_HIGH1	6,9559	0,0087
ING_LDH_LOW	6,6942	0,01
LAST_BLR_LOW	6,6851	0,0101
sedacion	6,517	0,0111
MAX_HCO3_LOW	6,5053	0,0111
I74	6,1998	0,0132
LAST_PH_HIGH2	6,1998	0,0132
MIN_BLR_HIGH2	6,1902	0,0133
MIN_PH_LOW2	5,7658	0,0168
MAX_ALP_HIGH2	5,211	0,023
I61	5,0964	0,0245
EX29	5,09	0,0246
MIN_PO2_MISS	5,0049	0,0258
MEAN_WBC_LOW2	4,9198	0,0271
vasoactivo	4,8673	0,0279
K65	4,7902	0,0292
MIN_HGB_LOW1	4,6811	0,0311
MIN_PO2_LOW1	4,3597	0,0374
LAST_AST_LOW	4,3547	0,0375
ING_PROT_LOW	4,2402	0,0401
C25	4,1017	0,0435
I21	4,1017	0,0435
I72	4,1017	0,0435
anticonv	4,059	0,0446
LAST_CRP_MISS	3,9265	0,0482
K80	3,9104	0,0487
ING_K_LOW	3,8915	0,0492
LAST_CRP_HIGH3	3,7011	0,0551
LAST_NA_LOW	3,3893	0,0664
LAST_CR_LOW	3,2866	0,0706
C22	3,2216	0,0734
MIN_AL_LOW	3,0883	0,0796
C71	3,0077	0,0836
MAX_AL_HIGH1	2,9925	0,0844
TIPO_QURG	2,9524	0,0865
ING_PCT_MISS	2,9388	0,0872
R57	2,9192	0,0883

Variable	Puntaje F	p-valor
ING_GLC_MISS	2,8588	0,0917
ING_UA_LOW	2,8156	0,0941
LAST_WBC_HIGH2	2,8144	0,0942
ING_PCO2_LOW	2,6767	0,1026
EX2	2,5772	0,1092
K83	2,497	0,1149
ING_GLC_HIGH3	2,4559	0,1179
MAX_PCO2_HIGH	2,4429	0,1188
TIPO_MED	2,4089	0,1214
EX4	2,2997	0,1302
EX13	2,19	0,1397
MAX_PLT_HIGH1	2,175	0,1411
MAX_BLR_HIGH1	2,175	0,1411
MAX_HGB_LOW2	2,1644	0,142
ING_BUN_HIGH1	2,1511	0,1433
I63	2,1183	0,1463
J15	2,1183	0,1463
MIN_GLC_HIGH1	2,0781	0,1502
ING_HCO3_HIGH	2,0685	0,1512
A40	2,0353	0,1545
B44	2,0353	0,1545
B57	2,0353	0,1545
B59	2,0353	0,1545
C03	2,0353	0,1545
C92	2,0353	0,1545
D33	2,0353	0,1545
D43	2,0353	0,1545
D65	2,0353	0,1545
D69	2,0353	0,1545
E11	2,0353	0,1545
I60	2,0353	0,1545
J86	2,0353	0,1545
K31	2,0353	0,1545
K46	2,0353	0,1545
K72	2,0353	0,1545
M32	2,0353	0,1545
M84	2,0353	0,1545
Z43	2,0353	0,1545
D32	1,9901	0,1591
MAX_PLT_HIGH2	1,9901	0,1591
EX26	1,8852	0,1705
MIN_HGB_MISS	1,6426	0,2007
ING_BUN_LOW	1,55	0,2139
K41	1,5474	0,2142

Variable	Puntaje F	p-valor
K75	1,5474	0,2142
K76	1,5474	0,2142
ING_CA_HIGH	1,5474	0,2142
MEAN_SO2_LOW3	1,5305	0,2168
MEAN_SO2_LOW2	1,5052	0,2206
A04	1,487	0,2234
C24	1,487	0,2234
C31	1,487	0,2234
J44	1,487	0,2234
K74	1,487	0,2234
N10	1,487	0,2234
N39	1,487	0,2234
EX10	1,4286	0,2327
ING_GLC_HIGH2	1,3403	0,2477
LOG_EDAD	1,2407	0,266
EX9	1,2209	0,2698
EX23	1,1425	0,2858
MIN_PO2_LOW2	1,1317	0,288
K81	1,1262	0,2892
S06	1,1262	0,2892

Fuente: Elaboración propia

A.2. Variables y puntaje F para corte p-valor=0.3 para modelo 48 horas

Tabla A.2: Puntaje F variables relevantes, modelo 48 horas

Variable	Puntaje F	p-valor
LAST_PH_LOW1	45,3689	6,431E-09
EX21	38,5438	1,468E-08
LAST_AL_HIGH2	35,5366	5,946E-08
LAST_AST_HIGH3	30,7136	5,769E-07
LAST_PLT_LOW1	30,0027	8,089E-07
EX14	21,7644	4,345E-06
LAST_LDH_HIGH	19,7662	1,166E-05
LAST_CRC_LOW1	14,7191	0,0001
LAST_K_HIGH	13,7726	0,0002
LAST_CR_HIGH2	13,756	0,0002
LOG_HOSP	13,2629	0,0003
LAST_BUN_HIGH2	13,0293	0,0003
MEAN_P_HIGH2	12,752	0,0004
LAST_BLR_HIGH2	11,6758	0,0007

Variable	Puntaje F	p-valor
VM	10,8273	0,0011
MIN_WBC_LOW1	10,3471	0,0014
LAST_AL_HIGH1	10,1991	0,0015
LAST_PCT_HIGH3	9,6979	0,002
MIN_WBC_HIGH1	9,6728	0,002
LAST_PCT_MISS	9,4343	0,0023
profilaxisgas	8,9862	0,0029
MIN_AST_LOW	8,7891	0,0032
ING_LDH_LOW	8,0576	0,0048
C16	8,0442	0,0048
EX5	7,7647	0,0056
MIN_PH_LOW2	7,3567	0,007
MAX_NA_HIGH	7,2916	0,0073
MEAN_UA_HIGH	7,098	0,0081
I74	6,8977	0,009
MIN_AL_LOW	6,6911	0,0101
MAX_ALP_HIGH1	6,5361	0,011
ING_GLC_LOW	6,3503	0,0122
LAST_CL_HIGH	6,2759	0,0127
MIN_CR_HIGH1	6,1617	0,0135
I61	5,8288	0,0163
MIN_HGB_LOW1	5,8216	0,0163
ING_ALP_HIGH2	5,7385	0,0171
MAX_WBC_LOW2	5,6383	0,0181
MAX_UA_LOW	5,6042	0,0184
MIN_WBC_HIGH2	5,3359	0,0215
LAST_CR_LOW	5,3033	0,0219
MIN_BLR_LOW	5,2063	0,0231
MIN_ALB_LOW1	4,9896	0,0261
MEAN_CRC_LOW2	4,9675	0,0264
C25	4,5566	0,0335
I21	4,5566	0,0335
I72	4,5566	0,0335
LAST_K_LOW	4,4559	0,0355
MAX_GLC_HIGH2	4,3071	0,0387
profilaxistrombo	4,2561	0,0398
EX29	4,1254	0,043
MAX_PCO2_LOW	4,1188	0,0431
vasoactivo	4,1121	0,0433
MIN_CL_LOW	3,9063	0,0489
C22	3,8863	0,0494
K65	3,7078	0,0549
LAST_COHG_MISS	3,7049	0,055
MAX_HCO3_LOW	3,6571	0,0566

Variable	Puntaje F	p-valor
MEAN_P_LOW	3,4992	0,0622
EX2	3,3971	0,0661
K80	3,3832	0,0667
MEAN_PCT_HIGH1	3,1721	0,0757
MAX_GLC_HIGH1	3,1702	0,0758
MAX_PCO2_HIGH	3,1536	0,0766
LAST_CRP_HIGH2	2,9751	0,0854
LAST_PLT_HIGH1	2,9128	0,0887
MAX_CRP_HIGH1	2,7545	0,0978
C71	2,7167	0,1002
I63	2,6664	0,1034
LAST_CRP_HIGH3	2,4706	0,1169
MIN_ALB_LOW2	2,4243	0,1203
sedacion	2,408	0,1216
MIN_SO2_LOW1	2,3205	0,1286
EX13	2,3162	0,1289
A40	2,2578	0,1338
B44	2,2578	0,1338
B57	2,2578	0,1338
B59	2,2578	0,1338
C03	2,2578	0,1338
C92	2,2578	0,1338
D33	2,2578	0,1338
D43	2,2578	0,1338
D65	2,2578	0,1338
D69	2,2578	0,1338
E11	2,2578	0,1338
I60	2,2578	0,1338
I85	2,2578	0,1338
J86	2,2578	0,1338
K31	2,2578	0,1338
K46	2,2578	0,1338
K72	2,2578	0,1338
M32	2,2578	0,1338
M84	2,2578	0,1338
Z43	2,2578	0,1338
LAST_CA_HIGH	2,2578	0,1338
K81	2,2547	0,1341
K83	2,2547	0,1341
LAST_CR_MISS	2,2072	0,1382
MIN_P_HIGH1	2,1089	0,1473
anticonv	2,0529	0,1528
MEAN_NA_LOW	2,0322	0,1549
LAST_PH_HIGH1	2,0114	0,157

Variable	Puntaje F	p-valor
MAX_CA_LOW	1,8953	0,1695
K41	1,829	0,1771
K75	1,829	0,1771
K76	1,829	0,1771
LAST_COHG_HIGH	1,829	0,1771
D32	1,7965	0,181
I62	1,7965	0,181
MIN_PLT_HIGH2	1,7965	0,181
MAX_BUN_LOW	1,7965	0,181
LAST_AST_HIGH2	1,7488	0,1869
TIPO_MED	1,7023	0,1928
MAX_BLR_HIGH1	1,6637	0,1979
MIN_PO2_LOW1	1,547	0,2144
MAX_PCT_HIGH2	1,547	0,2144
EX4	1,4959	0,2221
R57	1,4959	0,2221
MIN_HGB_LOW2	1,4851	0,2238
ING_HCO3_HIGH	1,4403	0,2309
LAST_BUN_HIGH1	1,4117	0,2356
A04	1,342	0,2474
C24	1,342	0,2474
C31	1,342	0,2474
J44	1,342	0,2474
N10	1,342	0,2474
N39	1,342	0,2474
MAX_PH_HIGH2	1,3213	0,2511
MIN_SO2_LOW2	1,1768	0,2787
MAX_CRC_LOW3	1,1732	0,2795
I71	1,1559	0,283
EX7	1,1275	0,289
TIPO_QURG	1,0978	0,2954

Fuente: Elaboración propia

A.3. Variables y puntaje F para corte p-valor=0.25 para modelo 72 horas

Tabla A.3: Puntaje F variables relevantes, modelo 72 horas

Variable	Puntaje F	p-valor
EX21	35,8847	5,746E-08
MEAN_PLT_LOW1	21,885	4,314E-06
MEAN_PH_LOW1	20,5697	8,201E-06

Variable	Puntaje F	p-valor
MAX_LDH_HIGH	18,982	1,791E-05
LAST_AL_HIGH2	17,6377	3,488E-05
MAX_AST_HIGH3	16,2038	7,141E-05
EX14	15,1634	0,0001
LAST_BLR_HIGH2	13,0051	0,0004
LOG_HOSP	11,895	0,0006
LAST_AL_HIGH1	11,7715	0,0007
MIN_HGB_LOW1	11,3553	0,0008
EX5	10,4398	0,0014
C16	9,368	0,0024
MIN_SO2_LOW1	9,2316	0,0026
MEAN_CR_HIGH2	8,354	0,0041
MAX_PH_HIGH2	8,2246	0,0044
MAX_UA_LOW	8,1106	0,0047
MEAN_K_HIGH	7,8337	0,0054
LAST_COHG_HIGH	7,8337	0,0054
MIN_AL_LOW	7,4636	0,0067
LAST_WBC_HIGH2	7,3247	0,0072
MIN_HGB_LOW2	7,3247	0,0072
ING_K_LOW	6,779	0,0097
MIN_CRC_LOW1	6,6948	0,0101
MAX_P_HIGH2	6,6792	0,0102
VM	6,2609	0,0129
ING_ALB_LOW1	6,1265	0,0138
ING_LDH_LOW	5,84	0,0162
LAST_PCT_HIGH3	5,7971	0,0166
MEAN_BUN_HIGH2	5,676	0,0178
LAST_WBC_HIGH1	5,5793	0,0188
LAST_PO2_LOW1	5,4621	0,0201
MEAN_PCO2_HIGH	5,312	0,0218
vasoactivo	5,1635	0,0237
C25	5,1624	0,0238
I72	5,1624	0,0238
I74	5,1624	0,0238
LAST_CA_HIGH	5,1624	0,0238
MAX_GLC_HIGH1	5,026	0,0257
LAST_PCT_MISS	4,897	0,0276
MIN_WBC_LOW1	4,4874	0,0349
I61	4,4007	0,0367
MAX_ALP_HIGH1	4,3521	0,0378
MIN_CRP_HIGH3	4,2494	0,0401
ING_ALP_HIGH2	4,1914	0,0415
TIPO_MED	3,9725	0,0471
ING_CRP_HIGH1	3,9301	0,0483

Variable	Puntaje F	p-valor
EX29	3,6678	0,0564
LAST_AST_LOW	3,6558	0,0568
MAX_NA_HIGH	3,5913	0,059
sedacion	3,5566	0,0602
ING_GLC_LOW	3,4334	0,0648
MIN_UA_HIGH	3,4334	0,0648
MAX_CA_LOW	3,4035	0,066
MAX_CR_HIGH1	3,2976	0,0703
MIN_PH_LOW2	3,2893	0,0707
EX13	3,2458	0,0726
MEAN_BUN_LOW	3,2453	0,0726
MEAN_CRP_HIGH2	3,2303	0,0733
MAX_PH_HIGH1	3,1823	0,0754
rnm	2,9586	0,0864
profilaxisgas	2,9187	0,0885
MEAN_PLT_HIGH1	2,9187	0,0885
K80	2,8355	0,0932
profilaxistrombo	2,7389	0,0989
A40	2,5519	0,1112
B44	2,5519	0,1112
B57	2,5519	0,1112
B59	2,5519	0,1112
C03	2,5519	0,1112
C48	2,5519	0,1112
D33	2,5519	0,1112
D43	2,5519	0,1112
D65	2,5519	0,1112
E11	2,5519	0,1112
I21	2,5519	0,1112
I85	2,5519	0,1112
J86	2,5519	0,1112
K31	2,5519	0,1112
K46	2,5519	0,1112
M32	2,5519	0,1112
Z43	2,5519	0,1112
C22	2,5302	0,1127
K65	2,5302	0,1127
ING_COHG_MISS	2,5028	0,1147
TIPO_QURG	2,4106	0,1215
LAST_HCO3_LOW	2,408	0,1217
LAST_AST_HIGH2	2,3698	0,1247
ING_BUN_HIGH1	2,2997	0,1304
ING_ALB_LOW3	2,257	0,134
K41	2,2061	0,1385

Variable	Puntaje F	p-valor
K76	2,2061	0,1385
LAST_CL_HIGH	2,0889	0,1494
LAST_GLC_HIGH3	2,0652	0,1517
K81	2,0008	0,1582
I63	1,9144	0,1675
EX20	1,9132	0,1676
MAX_NA_LOW	1,9104	0,1679
MIN_BLR_LOW	1,7665	0,1848
MEAN_CRC_LOW2	1,7579	0,1859
MAX_HCO3_HIGH	1,7506	0,1868
EX2	1,7427	0,1878
EX26	1,696	0,1938
I71	1,696	0,1938
R57	1,6618	0,1983
I62	1,5934	0,2078
K83	1,5934	0,2078
MEAN_PLT_HIGH2	1,5934	0,2078
EX24	1,4997	0,2216
J96	1,4243	0,2336
K57	1,4243	0,2336
MAX_PCO2_LOW	1,378	0,2413

Fuente: Elaboración propia

Apéndice B

Coeficientes Logit

En esta sección del apéndice se adjuntan las variables utilizadas para entrenar el modelo Logit, con su correspondiente coeficiente β .

B.1. Variables y coeficiente β para modelo 24 horas

Tabla B.1: Variables y coeficiente β para modelo 24 horas

Variable	β
Intercepto	-0.4740
TIPO_MED	-0,175
TIPO_QURG	0,456
LOG_EDAD	-0,7167
LOG_HOSP	0,3736
VM	0,7828
anticonv	-0,4322
profilaxisgas	-0,9262
sedacion	0,8196
vasoactivo	0,0466
EX10	-0,5633
EX13	0,3056
EX14	0,4193
EX2	0,2458
EX21	1,0116
EX23	-0,5325
EX26	-0,5962
EX29	-0,6587
EX4	-1,0492
EX5	0,5056
EX9	-0,2608

Variable	β
A04	-0,4629
A40	0,6911
B44	0,3302
B57	0,6192
B59	0,7613
C03	0,3352
C16	0,4732
C22	0,4708
C24	-0,5217
C25	0,2963
C31	-0,152
C71	-0,2934
C92	0,0198
D32	-0,3092
D33	0,3423
D43	0,3263
D65	0,0619
D69	0,5004
E11	0,1356
I21	1,0689
I60	0,7486
I61	1,5667
I63	1,7344
I72	0,4211
I74	0,5981
J15	1,0175
J44	-0,3515
J86	0,4529
K31	0,5683
K41	0,7687
K46	0,4789
K65	0,173
K72	0,0326
K74	-0,5934
K75	0,313
K76	0,1917
K80	-1,3079
K81	-0,5752
K83	-0,563
M32	0,3517
M84	0,0731
N10	-0,3779
N39	-0,2485
R57	0,7969

Variable	β
S06	0,0745
Z43	0,2005
MIN_CR_HIGH1	0,1693
LAST_CR_LOW	0,2325
LAST_CR_HIGH2	0,2415
MAX_CRC_LOW2	0,1311
MIN_CRC_LOW1	0,3215
MAX_AL_HIGH1	-0,2454
MIN_AL_LOW	-0,7968
LAST_AL_HIGH2	1,3373
MAX_NA_HIGH	0,0459
LAST_NA_LOW	-0,26
ING_K_LOW	-0,7236
MEAN_K_HIGH	0,183
MIN_PH_LOW2	-0,1368
LAST_PH_LOW1	0,5725
LAST_PH_HIGH2	1,3231
ING_PCO2_LOW	-0,1741
MAX_PCO2_HIGH	0,0112
ING_HCO3_HIGH	0,0721
MAX_HCO3_LOW	0,3468
MAX_HGB_LOW2	0,0331
MIN_HGB_LOW1	0,3551
MIN_HGB_MISS	0,277
MIN_WBC_LOW1	0,8102
MEAN_WBC_LOW2	0,7946
LAST_WBC_HIGH1	-0,2281
LAST_WBC_HIGH2	0,1707
MAX_PLT_HIGH1	0,0267
MAX_PLT_HIGH2	-0,2815
LAST_PLT_LOW1	0,2212
LAST_CRP_HIGH3	0,2423
LAST_CRP_MISS	-0,086
ING_GLC_LOW	0,3732
ING_GLC_HIGH2	-0,4668
ING_GLC_HIGH3	0,6597
ING_GLC_MISS	0,5145
MIN_GLC_HIGH1	0,5375
ING_BUN_LOW	-0,1182
ING_BUN_HIGH1	0,0508
LAST_BUN_HIGH2	0,4398
ING_UA_LOW	-0,4809
LAST_UA_HIGH	0,236
ING_CA_HIGH	0,1079

Variable	β
LAST_P_HIGH2	-0,3927
ING_PROT_LOW	0,1457
MIN_ALB_LOW1	0,6711
MAX_BLR_HIGH1	0,4044
MIN_BLR_HIGH2	0,1228
LAST_BLR_LOW	-0,665
MEAN_AST_HIGH3	0,4483
LAST_AST_LOW	-0,3867
MAX_ALP_HIGH2	0,0427
MEAN_ALP_HIGH1	0,1161
ING_LDH_LOW	-0,9713
ING_LDH_HIGH	0,455
MIN_PO2_LOW1	-0,0596
MIN_PO2_LOW2	-0,1154
MIN_PO2_MISS	0,6389
MIN_SO2_LOW1	0,1366
MEAN_SO2_LOW2	0,6341
MEAN_SO2_LOW3	0,1479
ING_PCT_HIGH3	-0,5071
ING_PCT_MISS	0,1706

Fuente: Elaboración propia

B.2. Variables y coeficiente β para modelo 48 horas

Tabla B.2: Variables y coeficiente β para modelo 48 horas

Variable	β
Intercepto	-0.9735
TIPO_MED	-0,5228
LOG_HOSP	0,3272
VM	0,5675
anticonv	0,0533
profilaxisgas	-0,9005
profilaxistrombo	-0,4225
sedacion	0,0859
vasoactivo	0,0524
EX13	-0,1432
EX14	0,2497
EX2	0,2294
EX21	0,9311
EX29	-0,7331
EX5	1,1649

Variable	β
A40	0,6073
B44	0,1693
B57	0,7625
B59	0,4793
C03	0,3597
C16	0,1392
C22	0,4408
C25	0,1452
C71	-0,3188
C92	0,0413
D32	-0,1969
D33	0,4753
D43	0,5059
D65	0,1271
D69	0,2622
E11	0,0939
I21	0,9774
I60	0,8359
I61	1,2388
I62	-0,3686
I63	1,5939
I72	0,3637
I74	0,8556
I85	0,3308
J86	0,1432
K31	0,5656
K41	0,446
K46	0,5994
K65	0,2903
K72	0,0009
K75	0,3545
K76	-0,0293
K80	-1,1096
K81	-0,1481
K83	-0,5554
M32	0,0661
M84	0,1226
Z43	0,0442
MIN_CR_HIGH1	0,2964
LAST_CR_LOW	0,1358
LAST_CR_HIGH2	0,2095
LAST_CR_MISS	1,0952
MEAN_CRC_LOW2	-0,3937
LAST_CRC_LOW1	-0,1275

Variable	β
MIN_AL_LOW	-0,8771
LAST_AL_HIGH1	0,9558
LAST_AL_HIGH2	1,0064
MAX_NA_HIGH	0,2364
MEAN_NA_LOW	-0,6227
LAST_K_LOW	0,9287
LAST_K_HIGH	0,4299
MIN_CL_LOW	0,8615
LAST_CL_HIGH	-0,8581
MIN_PH_LOW2	-0,065
LAST_PH_LOW1	1,121
LAST_PH_HIGH1	-0,0091
MAX_PCO2_LOW	-0,221
MAX_PCO2_HIGH	0,2237
MAX_HCO3_LOW	0,2965
MIN_HGB_LOW1	0,4157
MAX_WBC_LOW2	0,5662
MIN_WBC_LOW1	0,6774
MIN_WBC_HIGH1	-0,1847
MIN_WBC_HIGH2	0,5123
MIN_PLT_HIGH2	-0,57
LAST_PLT_LOW1	0,2907
LAST_PLT_HIGH1	0,3774
MAX_CRP_HIGH1	-0,7758
LAST_CRP_HIGH2	-0,8409
LAST_CRP_HIGH3	-0,5158
ING_GLC_LOW	0,1577
MAX_GLC_HIGH1	0,665
MAX_GLC_HIGH2	-0,426
MAX_BUN_LOW	-0,1365
LAST_BUN_HIGH2	0,4478
MAX_UA_LOW	-0,463
MEAN_UA_HIGH	0,1086
MAX_CA_LOW	-0,7107
LAST_CA_HIGH	0,2532
MIN_P_HIGH1	-0,4954
MEAN_P_LOW	-0,7323
MEAN_P_HIGH2	-0,3904
MIN_ALB_LOW1	0,6469
MIN_ALB_LOW2	0,1155
MAX_BLR_HIGH1	0,0482
MIN_BLR_LOW	-0,1028
LAST_BLR_HIGH2	0,0921
MIN_AST_LOW	-0,4562

Variable	β
LAST_AST_HIGH2	-0,8704
LAST_AST_HIGH3	0,5363
ING_ALP_HIGH2	0,0791
MAX_ALP_HIGH1	0,1675
ING_LDH_LOW	-0,9857
LAST_LDH_HIGH	0,3717
LAST_COHG_HIGH	0,4614
LAST_COHG_MISS	0,1655
MIN_SO2_LOW1	-0,4492
MEAN_PCT_HIGH1	0,8182
LAST_PCT_HIGH3	0,2053
LAST_PCT_MISS	-0,8327

Fuente: Elaboración propia

B.3. Variables y coeficiente β para modelo 72 horas

Tabla B.3: Variables y coeficiente β para modelo 72 horas

Variable	β
Intercepto	-1,6582
TIPO_MED	-0,4133
LOG_HOSP	0,35
VM	0,0824
vasoactivo	0,3357
EX14	0,627
EX21	1,3494
EX5	0,9656
C16	0,4742
C25	0,2361
I61	1,0675
I72	0,4093
I74	0,7919
MEAN_CR_HIGH2	0,2507
MIN_CRC_LOW1	-0,1414
MIN_AL_LOW	-1,169
LAST_AL_HIGH1	0,7161
LAST_AL_HIGH2	0,993
ING_K_LOW	-0,935
MEAN_K_HIGH	0,9211
MAX_PH_HIGH2	0,9604
MEAN_PH_LOW1	0,763
MEAN_PCO2_HIGH	0,5988

Variable	β
MIN_HGB_LOW1	-0,3938
MIN_HGB_LOW2	-0,9547
MIN_WBC_LOW1	0,7646
LAST_WBC_HIGH1	-0,8533
LAST_WBC_HIGH2	-0,0033
MEAN_PLT_LOW1	0,1231
ING_CRP_HIGH1	-0,6936
MIN_CRP_HIGH3	-0,0878
MAX_GLC_HIGH1	0,6017
MEAN_BUN_HIGH2	-0,0499
MAX_UA_LOW	-1,1388
LAST_CA_HIGH	0,6595
MAX_P_HIGH2	-0,0152
ING_ALB_LOW1	0,4544
LAST_BLR_HIGH2	-0,104
MAX_AST_HIGH3	0,2883
ING_ALP_HIGH2	0,4451
MAX_ALP_HIGH1	0,146
ING_LDH_LOW	-0,8053
MAX_LDH_HIGH	0,5898
LAST_COHG_HIGH	0,8537
LAST_PO2_LOW1	0,341
MIN_SO2_LOW1	-0,028
LAST_PCT_HIGH3	-0,3186
LAST_PCT_MISS	-0,423

Fuente: Elaboración propia