



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Trabajo Final: Diplomado en Inteligencia Artificial
Aplicada a Ciencias de la Salud

Título: Predicción de Hospitalización de pacientes de urgencia mediante un modelo de clasificación

Autora: Ignacia Pérez Arce



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Objetivo:

El objetivo del presente trabajo fue desarrollar un modelo de aprendizaje automático supervisado basado en **Regresión Logística** para predecir si un paciente admitido en un servicio de urgencia será **hospitalizado o dado de alta**, utilizando variables clínicas iniciales como edad, signos vitales y forma de ingreso. El propósito clínico de este modelo es **apoyar la gestión del flujo hospitalario**, anticipando la demanda de camas en unidades de urgencia y hospitalización.

Introducción:

Las Unidades de Urgencia Hospitalaria en Chile o servicios de urgencia (SU), son una de las principales puertas de entrada al sistema de salud (1). Es de conocimiento público que estos servicios suelen encontrarse saturados, con largas horas de espera para la atención sobretodo en temporadas de alta demanda como lo es la campaña de invierno. En este artículo de Radio Bío Bío (2), se describe una noticia de 2020 donde el SU del Hospital El Pino estaba colapsado en su capacidad.

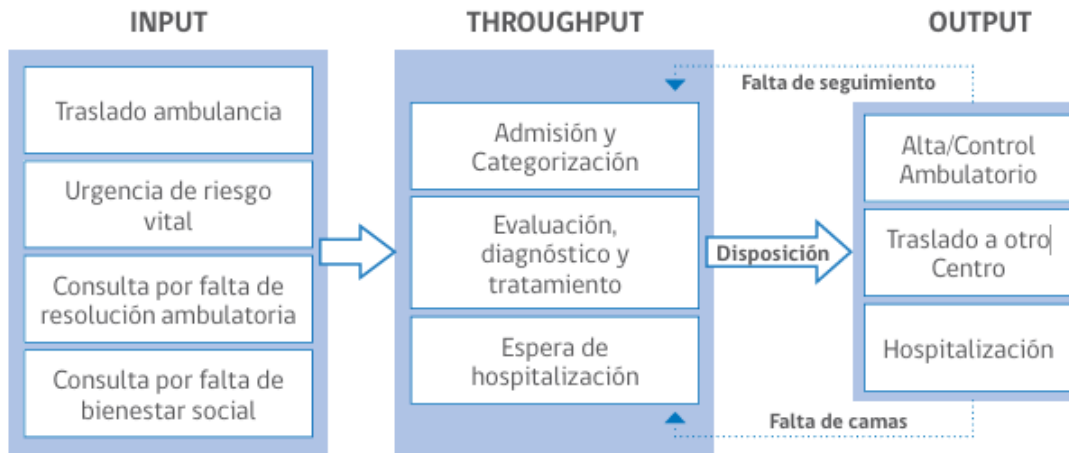
Si bien lo anteriormente señalado es conocido por la población general, existe un desconocimiento sobre la causa raíz de la saturación de los servicios de urgencia y esta radica en los pacientes hospitalizados en espera de cama de internación definitiva que deben ser encamados desde los SU. Con ello el recurso físico de camillas y boxes de atención debe ser destinado a pacientes que consultaron en estos servicios y que por falta de camas en la red pública deben mantenerse internados en el mismo servicio de urgencia.

Según las orientaciones técnicas para el rediseño al proceso de atención de urgencia de adulto en las unidades de urgencia hospitalaria (3), el modelo de atención se grafica de la siguiente forma:



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

FIGURA 1. MODELO DE ATENCIÓN DE URGENCIA



Fuente: Construido en base de "A conceptual model of emergency department crowding, Asplin Br, [Ann Emerg Med. 2003;42:173-180.]"

Donde el Throughput representa el accionar del personal de urgencia. La falta de output en los pacientes en espera de hospitalización por falta de camas es el principal responsable de la saturación de estos servicios.

Cabe cuestionarse entonces, ¿existirá una forma de predecir cuál de los pacientes que son admitidos en el SU deberán ser hospitalizados? Contar con esta información permitiría a los SU anticiparse a su demanda y evaluar medidas para articular la red de atención pública mediante Atención Primaria de Salud (APS) y Unidad de Gestión Centralizada de Camas (UGCC) con el objetivo de descongestionar el servicio y mantener un flujo de atención dinámico.

A nivel nacional el Departamento de Estadísticas e Información en Salud (DEIS) es el departamento de MINSAL que disponibiliza datos relacionados con la salud de nuestra población, para urgencias pone a disposición el portal <https://atencionesurgencia.minsal.cl/> (4), el cual permite visualizar las atenciones y hospitalizaciones diarias de urgencia. Sin embargo, este portal solo entrega el conteo de hospitalizaciones en urgencia según causa y semana estadística y no entrega datos respecto del estado del paciente que permitan establecer variables predictoras de hospitalización de los mismos.

Ante esta limitante y a que actualmente la autora de este trabajo no se encuentra trabajando en un área que le permita acceder a datos de pacientes (y que ante el supuesto hubiera realizado un uso responsable de los mismos ocultando datos que permitan identificar a estos pacientes). Es que decide construir este modelo en base



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

a un dataset disponible en la web a través del sitio Kaggle.com (5). Este dataset fue construido en base a pacientes adultos admitidos en 2 servicios de urgencia extranjeros en un periodo cercano a un año con un total de 1267 registros.

Y si bien es un dataset extranjero, las variables que se utilizan concuerdan con los datos que se registran cuando un paciente es atendido en un SU en Chile.

Por ello se opta en utilizar este dataset ya que puede ser extrapolable a la realidad de cualquier servicio de urgencia del país y a modo de ejercicio permitirá construir un modelo predictivo que no tenga el objetivo de reemplazar la ronda diaria que realiza el personal de urgencias para definir el destino de los pacientes que se encuentran en el servicio de urgencia, sino que sea un complemento a su ejercicio diario permitiendo entregar un análisis general de la situación de los pacientes en urgencia para anticiparse a la necesidad de camas y otras acciones necesarias para descongestionar el servicio y poder dar continuidad al proceso de atención.

Descripción de datos:

Como se señaló anteriormente el dataset fue descargado del sitio Kaggle.com y consta de 1267 registros de pacientes adultos admitidos en servicios de urgencia de Corea del Sur. Cabe destacar que este dataset fue construido en base a una investigación cuyo objetivo fue medir la precisión de la categorización de urgencias y causas de mistriage (error de categorización) en este país. (6)

Las variables contenidas en el dataset corresponden a las siguientes:

Variable	Tipo	Descripción
Group	Numérica	1 o 2 dependiendo de si el paciente fue atendido en un servicio de urgencia local o regional.
Sex	Numérica	1= femenino, 2 = masculino
Age	Numérica	Edad en años
Patients number per hour	Numérica	Cantidad de pacientes admitidos en una hora.
Arrival mode	Numérica	Modalidad de llegada a la urgencia 1= caminando, 2= ambulancia pública, 3= vehículo privado, 4= ambulancia privada, 5= vehículo policial, 6= silla de ruedas, 7= otros.
Injury	Numérica	1 = sin herida, 2 = con herida
Chief complain	Catórica	Motivo de consulta del paciente



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Mental	Numérica	Estado mental del paciente 1= alerta, 2= respuesta a la voz, 3= respuesta al dolor, 4= sin respuesta.
Pain	Numérica	0= sin dolor, 1= con dolor
NRS Pain	Numérica	En caso de dolor, valoración del mismo en escala de 1 mínimo al 9 dolor máximo.
SBP Systolic blood pressure	Numérica	Presión sistólica en mmHg
DP Diastolic blood pressure	Numérica	Presión diastólica en mmHg
HR Heart rate	Numérica	Frecuencia cardiaca expresada en latidos por minuto
RR respiratory rate	Numérica	Frecuencia respiratoria expresada en respiraciones por minuto
BT body temperature	Numérica	Temperatura corporal en grados Celsius
Saturation	Numérica	Saturación de oxígeno en valores de 75 a 100%
KTAS_RN	Numérica	Categorización según criticidad de la urgencia donde 1, 2 y 3 se consideran de gravedad y 4 y 5 pacientes menos graves que podrían ser atendidos en otro centro.
Diagnosis in ED	Categórica	Diagnóstico en texto libre, no se utilizan clasificaciones internacionales como CIE-10
Disposition	Numérica	Destino del paciente: 1= Discharge (alta), 2= admission to ward (hospitalización cama básica), 3= admission to ICU(hospitalización cama UCI), 4= AMA discharge (alta con derivación), 5= transfer (traslado), 6=death (muerte), 7=surgery (pabellón).
KTAS_expert	Numérica	Categorización según criticidad de la urgencia por un experto, donde 1, 2 y 3 se consideran de gravedad y 4 y 5 pacientes menos graves que podrían ser atendidos en otro centro.
Error_group	Numérica	Tipo de error en la categorización. 1= signos vitales, 2= examen físico, 3= psiquiátrico, 4= dolor, 5= estado mental, 6= enfermedad subyacente, 7= registros de otras urgencias, 8= on set, 9= otras.
Lenght of stay	Numérica	Estadía en urgencias expresada en minutos
KTAS duration	Numérica	Duración de la categorización en minutos



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

mistriage	Numérica	0= correcto, 1=over triage (mayor gravedad a la real), 2= under triage (menor gravedad a la real).
-----------	----------	--

El dataset fue sometido a una revisión exploratoria inicial para identificar valores nulos, outliers y tipos de variables. Posteriormente, se realizó limpieza, imputación de valores faltantes y codificación de variables categóricas mediante OneHotEncoder. También se creó una nueva variable clínica derivada: la Presión Arterial Media (MAP), obtenida a partir de los valores de presión sistólica y diastólica.

Métodos y modelos:

El presente estudio utilizó un enfoque de aprendizaje supervisado para la predicción binaria del destino del paciente en urgencias (Ingreso hospitalario = 1; Alta = 0).

El conjunto de datos fue dividido en dos subconjuntos mediante la función `train_test_split` de `scikit-learn`: un 80% para entrenamiento y un 20% para prueba, utilizando una semilla aleatoria fija (`random_state=42`) y estratificación sobre la variable objetivo para mantener la proporción de clases en ambos subconjuntos.

El preprocesamiento de los datos se realizó de manera automatizada mediante un `ColumnTransformer` que integró dos flujos principales:

Variables numéricas: fueron imputadas con la mediana para manejar valores faltantes y posteriormente estandarizadas con `StandardScaler` a media 0 y desviación estándar 1.

Las variables incluyeron: edad, presión arterial sistólica (SBP), presión arterial diastólica (DBP), frecuencia cardíaca (HR), frecuencia respiratoria (RR), temperatura (BT) y la presión arterial media (MAP), calculada con la fórmula clínica $MAP = (SBP + 2 \times DBP) / 3$.

Variables categóricas: fueron imputadas con el valor más frecuente y transformadas mediante One Hot Encoding, generando variables binarias representativas.

Se desarrollaron y compararon dos modelos predictivos:

Regresión Logística (RL): seleccionada como modelo principal por su interpretabilidad y capacidad de estimar la influencia individual de cada predictor sobre la probabilidad de ingreso hospitalario. Se configuró con regularización L2, `max_iter=1000`, y `class_weight='balanced'` para compensar el leve desbalance de clases.



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

Random Forest (RF): implementado como modelo comparativo por su capacidad de capturar relaciones no lineales y manejar interacciones entre variables. Se configuró con 200 árboles ($n_estimators=200$), $class_weight='balanced'$, y $random_state=42$.

Ambos modelos se entrenaron sobre el mismo conjunto preprocesado, asegurando reproducibilidad y comparación directa.

Posteriormente, se realizó una búsqueda de hiperparámetros mediante GridSearchCV (validación cruzada de 5 pliegues) para optimizar la regularización en RL y la profundidad y número de árboles en RF.

Finalmente, se evaluaron los modelos en el conjunto de prueba. En el caso de la Regresión Logística, además se calcularon los odds ratio (OR) para interpretar la magnitud del efecto de cada variable sobre la probabilidad de hospitalización.

Evaluación del Modelo:

El rendimiento de ambos modelos se evaluó con las métricas estándar de clasificación:

exactitud (accuracy), precisión (precision), sensibilidad (recall) y puntaje F1 (F1-score), priorizando clínicamente la sensibilidad de la clase “Ingreso hospitalario”, dado que el objetivo principal es reducir falsos negativos (pacientes que requieren hospitalización pero son clasificados como alta).

El modelo de Regresión Logística alcanzó una exactitud del 74% y un recall del 70%, mientras que el Random Forest obtuvo una exactitud de 70% y un recall de 46%. Aunque ambos modelos mostraron precisión similar ($\sim 0.59-0.60$), la Logística demostró mejor capacidad de detección de ingresos reales, lo que es prioritario desde el punto de vista clínico.

El puntaje F1 de la clase ingreso fue de 0.65 para la Regresión Logística y 0.52 para el Random Forest, confirmando que el modelo lineal equilibra mejor precisión y sensibilidad.

Además, la matriz de confusión del modelo logístico evidenció una mayor proporción de verdaderos positivos y una reducción significativa de falsos negativos.

Para la Regresión Logística se construyeron también las curvas de precisión–recall y ROC, observándose un área bajo la curva (AUC) satisfactoria y un punto de equilibrio adecuado entre sensibilidad y precisión.

Se ajustó el umbral de decisión (threshold) bajo 0.5 para aumentar la sensibilidad por sobre el 80%, asumiendo un pequeño aumento en falsos positivos, lo cual es clínicamente aceptable en un escenario de urgencias.

Interpretación de resultados:

Los coeficientes y odds ratio obtenidos en la Regresión Logística permiten una interpretación clínica directa y coherente con la práctica asistencial.

Las variables con mayor impacto positivo sobre la probabilidad de ingreso fueron:

Modo de llegada en ambulancia privada ($OR \approx 2.35$) y ambulancia pública ($OR \approx 1.60$), lo que refleja que los pacientes trasladados suelen tener mayor gravedad.

Mayor edad ($OR \approx 1.42$), consistente con la vulnerabilidad clínica del adulto mayor.

Menor presión arterial sistólica y diastólica, asociadas con cuadros de shock o hipotensión, predictores clásicos de severidad.

Temperatura corporal elevada (BT) y alteración del nivel de conciencia (Mental 2 o 3), también asociados a mayor riesgo clínico y necesidad de hospitalización.

Por su parte, el Random Forest mostró una distribución de importancia de variables coherente con los hallazgos de la Logística, aunque sin una mejora sustancial del rendimiento.

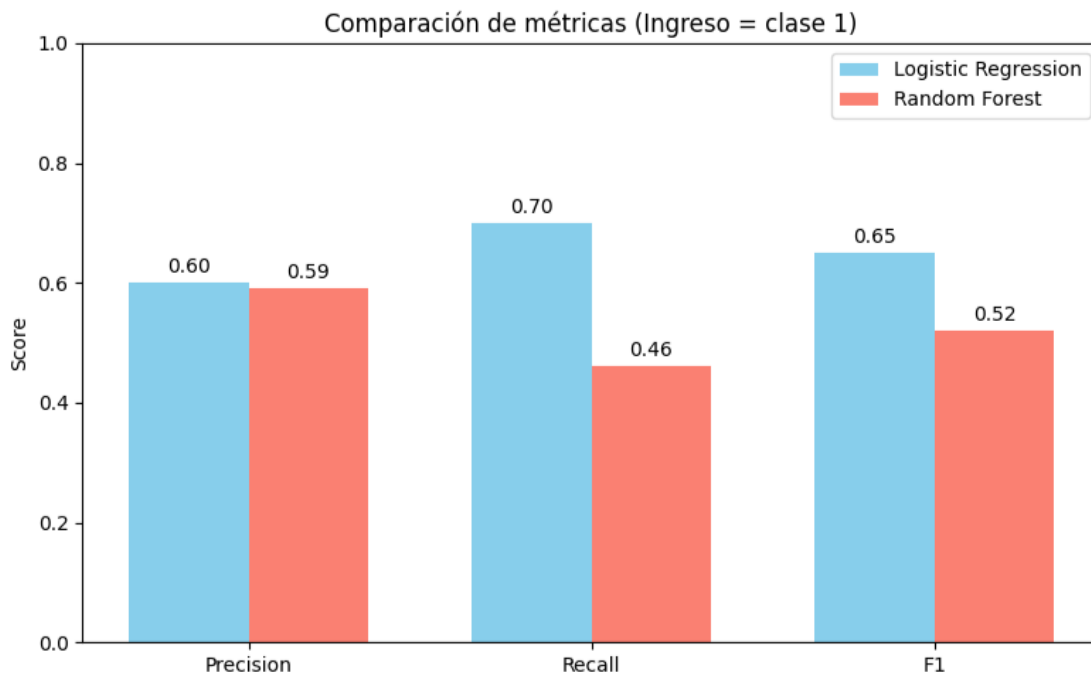
El análisis de importancia confirmó que las variables de modo de llegada, signos vitales y clasificación KTAS fueron las más determinantes.

Desde la perspectiva clínica, estos resultados son interpretables y reproducibles, lo que respalda la decisión de privilegiar la Regresión Logística como modelo final, dado que ofrece un equilibrio adecuado entre desempeño y transparencia.

Esto permite comunicar sus resultados a equipos médicos y de gestión sin perder rigor técnico.



UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA



Conclusiones:

Los resultados indican que el modelo de Regresión Logística constituye una herramienta prometedora para apoyar la decisión de hospitalización en servicios de urgencia, permitiendo identificar con buena sensibilidad a los pacientes con mayor riesgo de requerir cama hospitalaria.

El modelo de Random Forest fue útil como comparación metodológica, confirmando la robustez del enfoque y la coherencia de los predictores clínicos relevantes.

A futuro, se recomienda replicar este análisis con datos nacionales y evaluar su integración en sistemas de apoyo a la decisión clínica (CDSS), considerando variables organizacionales y temporales (p. ej., tiempos de espera o disponibilidad de camas).

Bibliografía:

1. Centro de Políticas Públicas, Universidad de los Andes. (2025). *Urgencias en Chile: Ante la demanda creciente, ¿Más servicios o mejor estrategia?*
<https://www.uandes.cl/wp-content/uploads/2025/01/Urgencias-en-Chile-Ante-la-demanda-creciente-¿Mas-servicios-o-mejor-estrategia.pdf>



UNIVERSIDAD TÉCNICA
FEDERICO SANTA MARÍA

2. BioBioChile. (2020, 20 de mayo). *Hospitales de la Región Metropolitana al borde del colapso: Urgencias saturadas y largas esperas*.
<https://www.biobiochile.cl/noticias/nacional/region-metropolitana/2020/05/20/hospitales-la-region-metropolitana-al-borde-del-colapso-urgencias-saturadas-largas-esperas.shtml>
3. Ministerio de Salud de Chile. (2019). *Atención de Urgencias en la Red Asistencial*. Biblioteca Virtual en Salud Chile.
https://docs.bvsalud.org/biblioref/minsai_chile/1/3/5131.pdf
4. Ministerio de Salud de Chile. (s. f.). *Atenciones de Urgencia en Chile*.
<https://atencionesurgencia.minsal.cl/>
5. Yıldız, İ. (2022). *Emergency Service Triage Application [Conjunto de datos]*. Kaggle. <https://www.kaggle.com/datasets/ilkeryildiz/emergency-service-triage-application/data>
6. Sun-Hee Moon, et. al. (2019). *Triage accuracy and causes of mistriage using the Korean Triage and Acuity Scale*. *PLOS ONE*, 14(6), e0216972.
<https://doi.org/10.1371/journal.pone.0216972>