

Jonathan Moya Carvajal¹ Cristian Martínez²
¹Magíster en Ciencias de la Computación, Universidad Católica del Maule,

²LITRP, Facultad de Ciencias de la Ingeniería, Universidad Católica del Maule,

Avenida San Miguel 3605, Talca, Chile

Introducción

El presente trabajo corresponde a un análisis detallado de registros de atención de pacientes que asistieron a la Unidad de Emergencias del Hospital San Juan de Dios de Curicó entre el 1 de Enero de 2014 y 1 Enero de 2015. Dichos registros corresponden a consultas de 139.704 pacientes. Nuestro objetivo al desarrollar esta investigación es poder identificar características en los datos que nos permitan agrupar pacientes con comportamiento similar con el propósito de predecir a posterior, cuestiones relacionadas con una mejor atención de los mismos. Para ello se aplicaron técnicas de Aprendizaje Automático. Mediante técnicas de Aprendizaje No Supervisado se desarrollaron modelos basados en K-Means que permitieron caracterizar diferentes grupos de pacientes. A su vez con técnicas de Aprendizaje Supervisado, se propusieron modelos basados en Random Forests, Support Vector Machines y XGBoost para predecir niveles de gravedad de la urgencia y necesidades de hospitalización.

Objetivos

El objetivo principal fue desarrollar modelos basados en Aprendizaje Automático para predecir nivel de gravedad y necesidad de hospitalización de pacientes de urgencias en etapas tempranas de atención. A su vez, los objetivos específicos son:

- Abordar un problema de interés para nuestra Comunidad.
- Aplicar técnicas y desarrollar modelos robustos y confiables de Aprendizaje Automático y Ciencia de Datos.
- Ofrecer una solución que sirva de base para una mejora del sistema hospitalario.

Metodología

Graphical Abstract

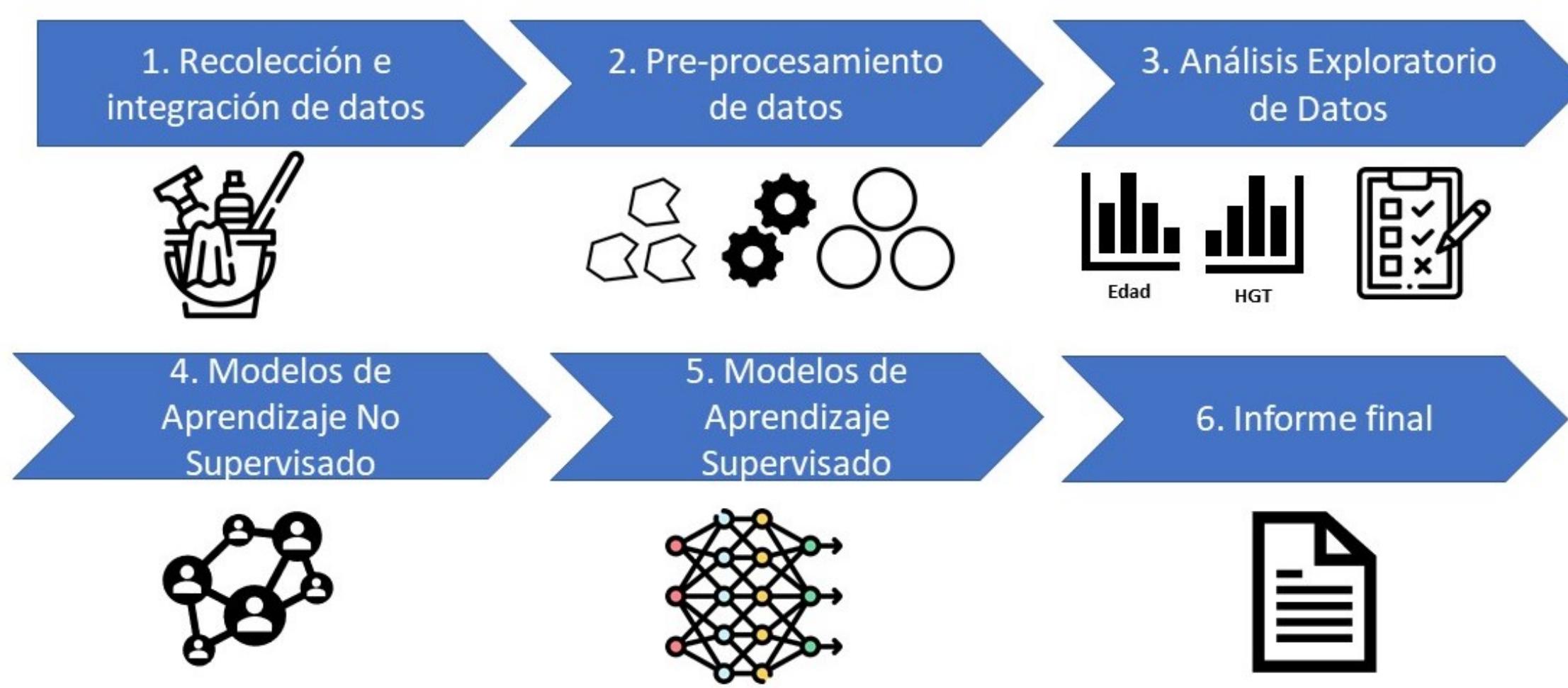


Figura 1. Flujo de trabajo propuesto para la presente investigación. Las actividades indicadas son iterativas y evolutivas hasta alcanzar resultados competentes.

Desarrollo del caso

Análisis Exploratorio de Datos

Se realizó una limpieza de datos y un posterior análisis exploratorio de la base de datos en estudio. La Figura 2 representa la relación entre la variable Nivel, que representa el nivel de gravedad asignado al paciente en el Triage con otras variables como: Pronóstico, que categoriza a priori la gravedad del diagnóstico, Destino, que representa el destino post atención del paciente y Hospitalizado, que indica si el paciente fue hospitalizado al final de su proceso de atención.

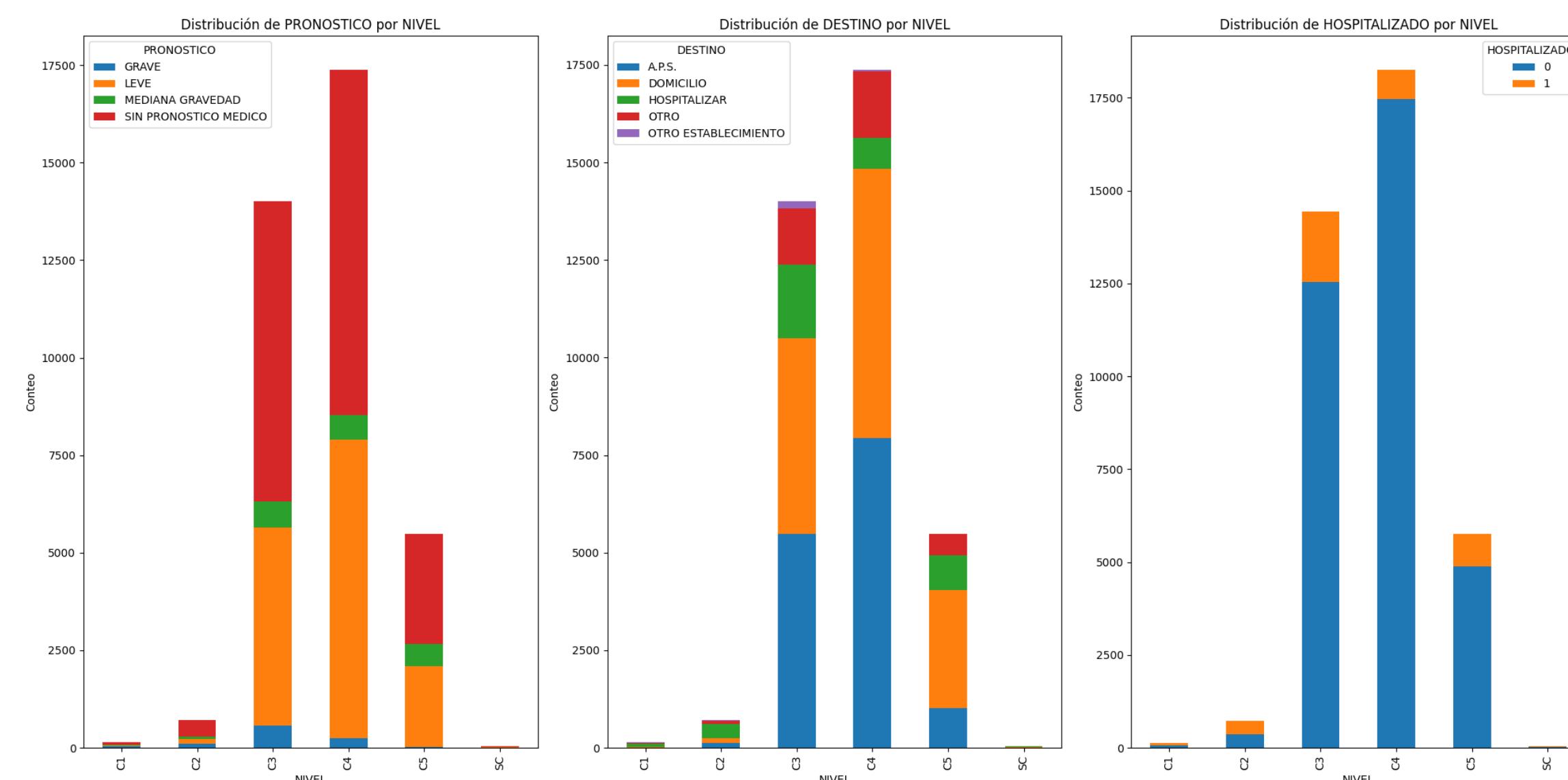


Figura 2. Relación entre cada una de las variables categóricas y el Nivel de gravedad

Modelos de aprendizaje No supervisado

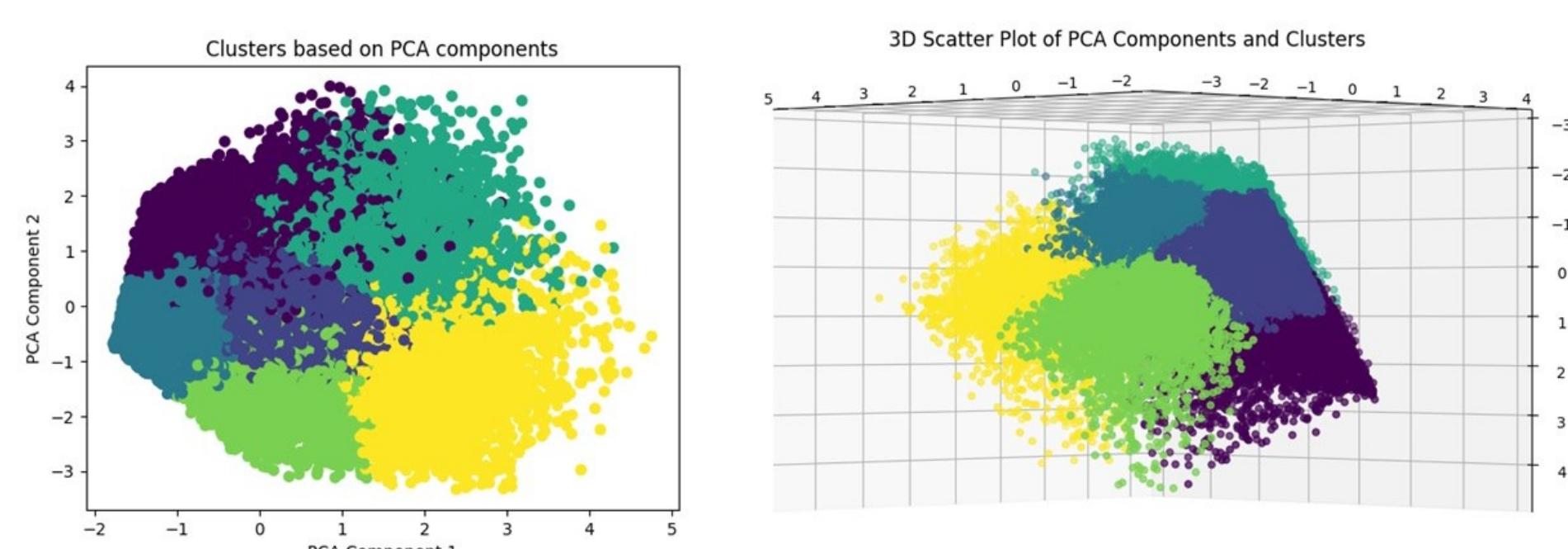


Figura 3. Representación gráfica en 2 y 3 dimensiones de los datos, luego de aplicar K-Means

Al clusterizar los datos de pacientes se generan 6 grupos los cuales son representados en la Figura 3. La composición de cada grupo presenta una proporción similar de cada una de las categorías vistas en la base de datos.

Desarrollo del caso

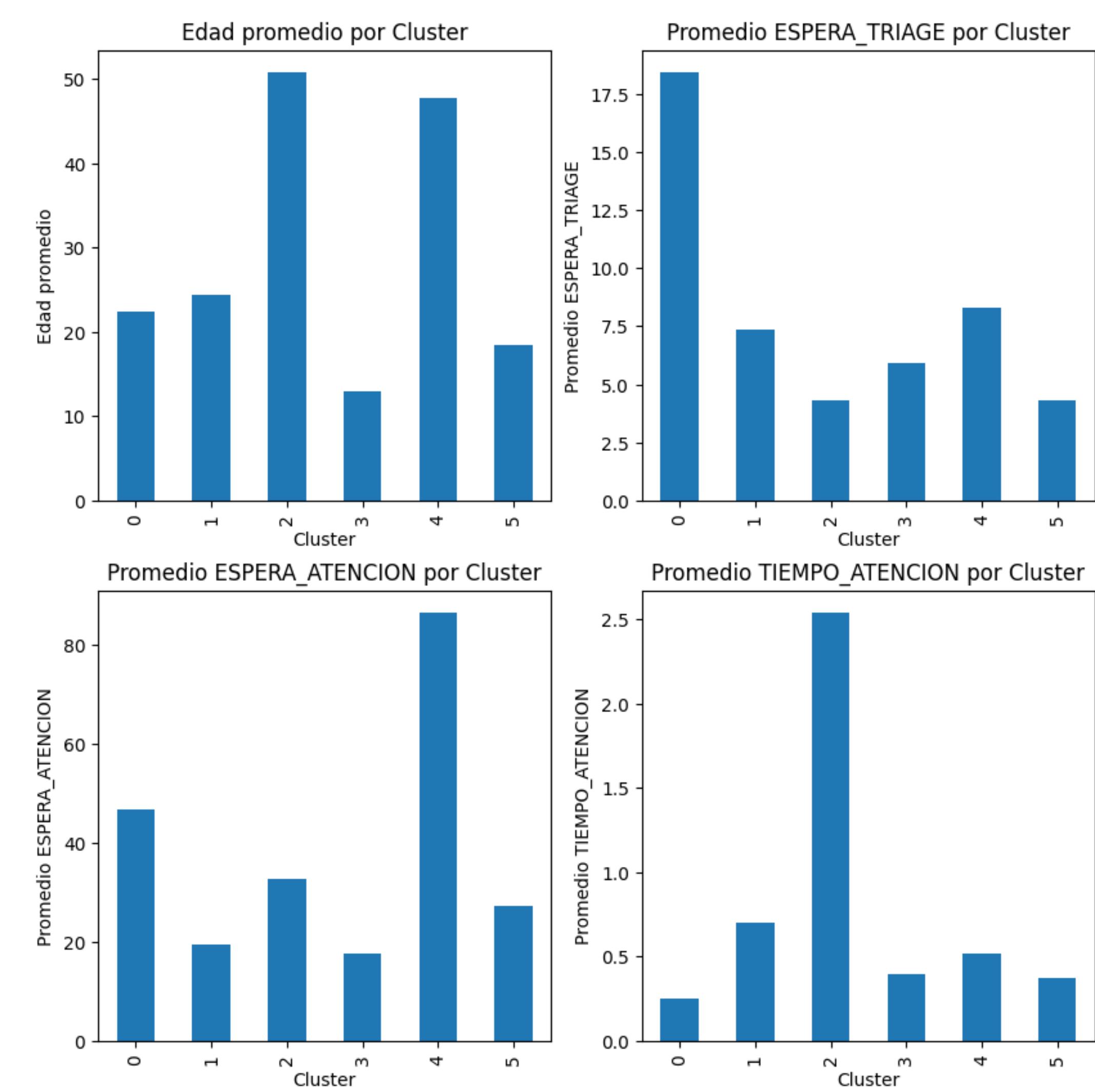


Figura 4. Relación entre cada una de las variables de tiempo y el clúster asignado

La Figura 4 presenta distintas variables de tiempo asociadas a las esperas experimentadas por los pacientes. Las imágenes muestran diferencias en los tiempos promedios, por los elementos de cada grupo, es decir, hay grupos que en promedio esperaron más que otros.

Resultado

Modelos de aprendizaje Supervisado



Figura 5. Resultados de predicción de tiempo de espera por atención

Se dividió la base de datos y se entrenaron modelos de machine learning con los elementos de cada cluster, de forma independiente para identificar diferencias en el nivel de precisión de los modelos.

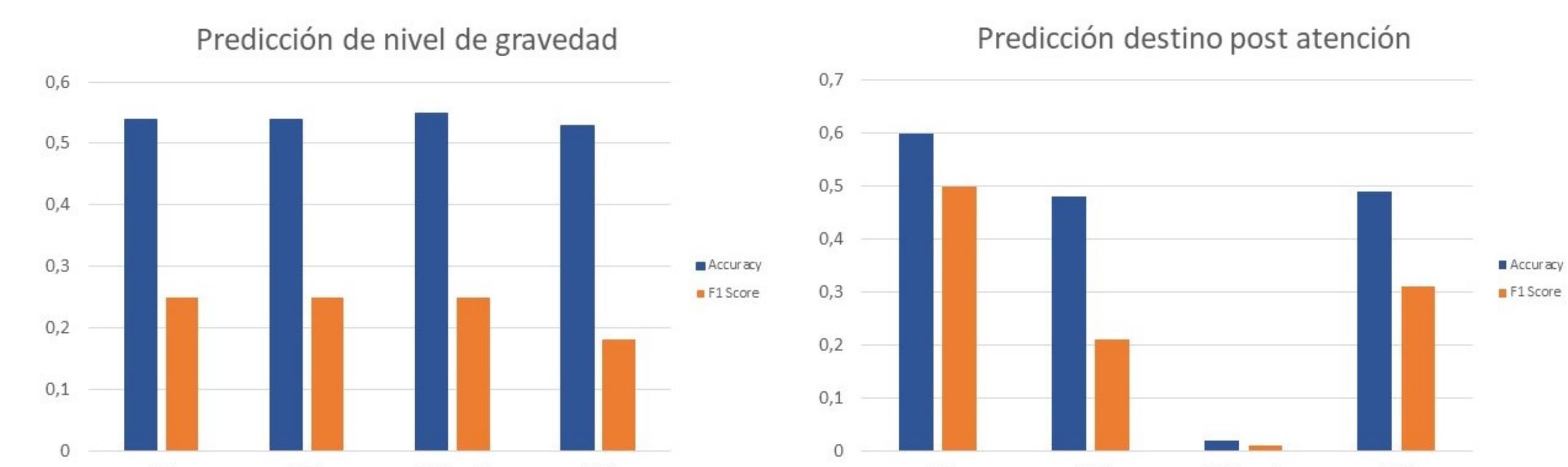


Figura 6. Predicción de nivel de gravedad y destino post atención

Se predijeron también las categorías: Nivel de gravedad y destino post atención, obteniendo resultados similares para cada clúster.

Discusión

Los modelos de aprendizaje automático, aplicados a los datos de Hospital San Juan de Dios de Curicó, han sido útiles para la identificación de características y patrones en los datos de pacientes. A través de la clusterización de los datos, fue posible agrupar pacientes con comportamientos similares, revelando diferencias en los tiempos de espera, que se tradujeron eficiencia de las predicciones asociadas a esa variable. Observamos también que al aplicar modelos de ML de manera segmentada, por cada cluster, se obtiene una mejor precisión pero con menor error que al entrenar con el conjunto de datos completo. Los modelos de predicción supervisada mostraron un desempeño aceptable, considerando que fueron entrenados con muy poca información.

Referencias

[1] Jonathan Moya-Carvajal, Francisco Pérez-Galarce, Carla Taramasco, César A. Astudillo, and Alfredo Candia-Véjar. MI models for severity classification and length-of-stay forecasting in emergency units. *Expert Systems with Applications*, 223:119864, 2023.