

Resumen del Proyecto Final

Machine Learning II

Predicción temprana de progresión a riesgo cardiovascular para priorización preventiva en APS rural.

Claudio Cárdenas Mansilla¹
Evelyn Sánchez Cabezas²

**Subdepartamento de Inteligencia Sanitaria, Dirección de Servicio de Salud Chiloé
Coordinación de Centros de Atención a la Comunidad, Facultad de Salud y Ciencias Sociales, Universidad de Las Américas**

En este proyecto se desarrolló un modelo de aprendizaje supervisado orientado a la predicción temprana del riesgo de progresión hacia criterios del Programa de Salud Cardiovascular (PSCV) en usuarios atendidos en Atención Primaria de Salud (APS), utilizando registros clínicos rutinarios provenientes del Examen de Medicina Preventiva del Adulto (EMPA) y controles cardiovasculares de la comuna de Quellón.

El problema se abordó como una tarea de clasificación binaria, donde la variable objetivo indica si un usuario progresó o no a criterios PSCV. El conjunto de datos incluyó variables clínicas y antropométricas básicas, tales como edad, sexo, presión arterial, circunferencia de cintura, colesterol total, peso y talla. Se realizó un análisis exploratorio de datos y un proceso de preprocesamiento que consideró tratamiento de valores atípicos, escalamiento de variables numéricas y codificación de variables categóricas, implementado mediante pipelines reproducibles.

Se entrenaron y compararon distintos modelos de aprendizaje supervisado, incluyendo regresión logística (en diversas variantes), KNN, SVM y XGBoost, utilizando validación cruzada estratificada y ajuste sistemático de hiperparámetros mediante Grid Search.

La métrica principal de evaluación fue PR-AUC, priorizando la correcta identificación de usuarios en riesgo. Los resultados evidenciaron que XGBoost optimizado presentó el mejor desempeño global, manteniendo un equilibrio adecuado entre sensibilidad, precisión y utilidad operativa en el contexto de APS. La regresión logística L1 mostró un desempeño comparable, destacando por su mayor interpretabilidad clínica.

Adicionalmente, se incorporó análisis de interpretabilidad mediante SHAP, permitiendo identificar variables clínicamente relevantes como edad, circunferencia de cintura y presión arterial sistólica.

Finalmente, se discutieron las principales limitaciones del modelo y se propuso un esquema conceptual de despliegue orientado a la priorización preventiva en APS, enfatizando su uso como herramienta de apoyo a la toma de decisiones clínicas y no como sistema de diagnóstico automático.