



Determinación de variables predictoras para tratamiento curativo con radioterapia en pacientes con adenocarcinoma de próstata en una cohorte del mundo real utilizando modelos de aprendizaje de máquina

»Luis Pino^{1,2,3}
»Juan Galvis⁴
»Margarita Hidalgo⁵
»María Maldonado⁴
»Iván Triana³
»Camilo Giedelmann⁶

¹ Centro de pensamiento Aipocrates

² OxLER SAS

³ Fundación Santa Fe de Bogotá

⁴ Clínica Los Nogales

⁵ Universidad del Rosario

⁶ Clínica de Marly

Introducción: el objetivo de este estudio es identificar las variables predictoras para el tratamiento curativo con radioterapia en pacientes con adenocarcinoma de próstata remitidos a una institución de referencia en Colombia utilizando dos modelos de regresión mediante aprendizaje de máquina (machine learning).

Materiales y métodos: se obtuvieron datos a partir de 208 historias clínicas de pacientes que recibieron radioterapia curativa para adenocarcinoma de próstata. Esta base de datos compiló un total de 40 variables agrupadas en características sociodemográficas, características clínicas como niveles de PSA previo a radioterapia y niveles iniciales de PSA.

Se determinó como variable objetivo la realización o no de radioterapia con intención curativa y se ejecutaron los siguientes pasos para la creación de dos modelos de predicción:

1. Análisis univariado, bivariado y multivariado (análisis correlacional simple),
 2. Preprocesamiento de datos,
 3. Ingeniería de variables para tratamiento de datos extremos, datos perdidos e imbalances,
 4. Creación de variables codificadas (one-hot encoder) y dummy para modelación,
 5. Desarrollo de dos modelos predictivos,
 - 5.1. Modelo de regresión logística múltiple (RLM),
 - 5.2. Modelo de árbol de decisión para regresión (CART),
 6. Ajuste de hiperparámetros,
 7. Medición comparativa de desempeño y
 8. Determinación de variables predictoras.
- Estos pasos se realizaron en el software Python versión 3.3. con el uso de librerías Matplotlib, Numpy, Pandas y Sci-Kit learn, entre otras.

Resultados:

Hallazgos descriptivos: los rangos de edad de

Autor para correspondencia: Pedro Ramos

Correo electrónico: pramos152@yahoo.com

Sociedad Colombiana de Hematología y Oncología. Este es un artículo Open Access bajo la licencia CC BY-NC-ND. (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

los pacientes que llegaron a radioterapia con intención curativa oscilaron entre 55 y 84 años con media de 71 años. La estratificación de riesgo según D'Amico fueron bajo riesgo 14 pacientes (14%), riesgo intermedio 52 pacientes (52.5 %) y alto riesgo 34 pacientes (34 %). Dos de los pacientes se clasificaron como metastásicos y se retiraron del análisis por ser outliers. El estado clínico más común a la remisión fue cT1 o cT2 en 80 pacientes (80 %). El tratamiento recibido fue en 74 % radioterapia más TDA la cual se extendió desde 6 a 24 meses según cada caso. Hubo cinco pacientes que tras arribar a radioterapia fueron dejados sin intervención, en vigilancia activa.

Modelos: se determinaron las siguientes variables para el análisis: edad, antecedentes familiares de cáncer de próstata, escala de riesgo, ISUP, iPSA, TNM, niveles de PSA previo a radioterapia, tratamiento inicial, cirugía, número de nodos resecaados, número de nodos comprometidos y tiempo a radioterapia.

Modelo de Regresión Logística Simple:

Error residual estándar 7.066

R2 0.63

R2 ajustado-0.126

F1 0.83

La Figura 1 muestra las variables predictoras por importancia. Los coeficientes de regresión más importantes fueron para las variables estadio T3, PSA preradioterapia igual o superior a 2.2 y

Gleason 5+4.

Modelo de CART: se construyó un modelo bajo criterio de clasificación Gini con estado de random=1 iniciando con 104 nodos, logrando una exactitud de 0.53 y un F1 de 0.65. Posteriormente se hizo ajuste de hiperparámetros utilizando GridSearch CV obteniendo 52 nodos. Posterior a este ajuste se obtuvo una exactitud final de 0.76 con un score F1 de 0.83 que se consideró como generalizable. Las variables finales del modelo CART se presentan en la Figura 2.

Conclusiones: como puede observarse, el modelo de RL múltiple logra un desempeño medido por exactitud que es inferior al desempeño logrado por el modelo CART. Con base en el modelo CART se definieron 6 variables predictoras que son las de mayor peso: niveles de PSA previo a radioterapia, estadio TNM (estadio T3), iPSA, escala de riesgo (intermedio/alto), antecedentes familiares de cáncer de próstata y escala de Gleason ($\geq 3+4$). Este modelo plantea la base para generar una escala local que pueda predecir los pacientes con adenocarcinoma de próstata con intención curativa que son potenciales candidatos a tratamiento con radioterapia.

Palabras clave:

Próstata; machine learning; variables predictoras; curación.

Figuras y Tablas:

Figura 1.

Variables predictoras por importancia

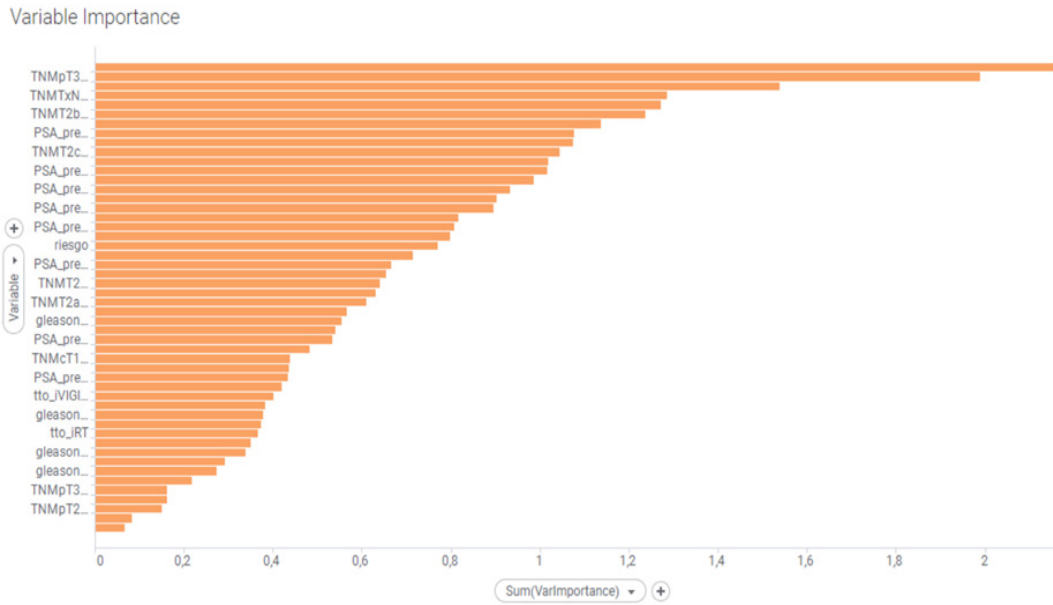


Figura 2.

Variables finales del modelo CART

