

GENERALIDADES DEL MACHINE LEARNING Y SU APLICACIÓN EN LA GESTIÓN SANITARIA EN SERVICIOS DE URGENCIA

Zuleika Damaris Ocampo Moreira

19 de octubre de 2023

Resumen

Este artículo discute la aplicación de técnicas de aprendizaje automático en la resolución de problemas del mundo real. Exploramos varios algoritmos y metodologías y presentamos resultados experimentales para demostrar su eficacia.

1. Introducción

Los procesos involucrados en la atención de salud generan una gran cantidad de información que resulta difícil analizar. Esto se debe, entre otras cosas, al volumen, velocidad de producción y diversidad de los datos (por ejemplo: texto, imágenes, tiempos, registros administrativos).

2. Una aproximación general al aprendizaje automático o Machine Learning

El ML surge en la década de los cincuenta como un recurso para emular, computacionalmente, elementos del proceso cognitivo humano a través de reconociendo de patrones y procesos de toma de decisión. En el ámbito de la medicina ha sido utilizado para aumentar la precisión diagnóstica, hacer predicciones de mortalidad hospitalaria o predecir la necesidad de ciertas terapias.

Los algoritmos de ML pueden ser clasificados en: supervisados, no supervisados y parcialmente supervisados

3. Algoritmos no supervisados

Los algoritmos no supervisados se caracterizan por su capacidad para extraer, clasificar y reducir la dimensionalidad. Este tipo de análisis es útil para la gene-

ración de nuevas hipótesis de trabajo. Algunos ejemplos de este tipo algoritmos son: Cluster Analysis (o Análisis de Conglomerados) y Análisis de Componentes Principales (PCA, en inglés).

4. Algoritmos supervisados

El objetivo de los algoritmos supervisados es la predicción. Para esto, el algoritmo es entrenado para relacionar un conjunto de variables con un resultado de interés. Una vez que el entrenamiento se ha completado, el algoritmo puede hacer predicciones utilizando nuevos datos.

5. Algoritmos parcialmente supervisados

Los algoritmos parcialmente supervisados comparten características de aquellos supervisados y no supervisados. Una de sus ventajas es poder trabajar con datos en donde no todas las unidades cuentan con un valor en la variable resultado. Por ejemplo, un algoritmo supervisado puede ser entrenado con un subconjunto de los datos disponibles y luego aplicado para predecir los valores en la variable resultado en aquellos casos que no contaban con este valor.

6. Utilidad del Machine Learning en la gestión de salud

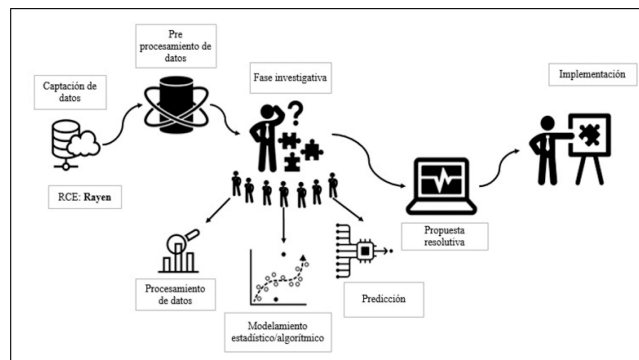


Figura 1: Modelo de Gestión del Conocimiento, Unidad de Ciencia de Datos del Servicio de Salud Metropolitano Sur. Autor: Enf. Natalia Gillies, Subjefa Departamento Gestión TIC SSMS.

Los algoritmos de aprendizaje automático, además de ser útiles para realizar predicciones clínicas y epidemiológicas, también pueden ser aplicados en la gestión de servicios de salud. En esta área, tienen el potencial de aportar al análisis

de los procesos clínicos y administrativos que tienen lugar en una organización de salud. Un proceso puede ser definido como una serie de pasos interconectados dentro de un contexto organizacional cuyo objetivo es crear bienes o servicios. Algunos ejemplos de procesos son la atención de un paciente en el servicio de urgencia, la gestión de insumos, la organización del personal sanitario, entre otros.

7. Aplicaciones en el triage

El triage es proceso clasificación de riesgo que permite discriminar entre quienes requieren atención inmediata en el servicio de urgencia de quienes pueden esperar más tiempo. Al clasificar a los usuarios los profesionales de salud deben lidiar con al menos dos problemas subestimar o sobre estimar el riesgo, es decir, asignar una mayor prioridad de atención a una persona con un situación de bajo riesgo vital o clasificar con un bajo riesgo a alguien que posee una condición crítica de salud.

8. Identificación de pacientes policonsultantes

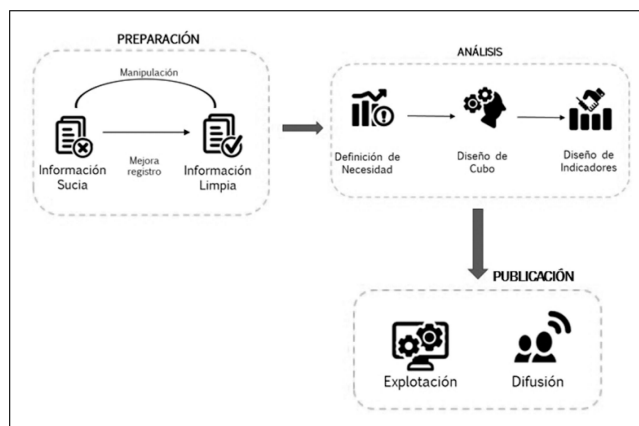


Figura 2: Esquema general de las etapas de la aplicación de ML para detectar usuarios potencialmente policonsultantes en servicios de urgencia. Autor: Elaboración propia.

En los servicios de urgencia los pacientes policonsultantes o hiperfrecuentadores son responsables de hasta un 30 por ciento del total de consultas en estos servicios. El fenómeno de la policonsulta puede ser explicado tanto por variables propias de los usuarios (por ejemplo, la edad de los usuarios, el grado de vulnerabilidad, la presencia de patologías psiquiátricas, entre otros) como del sistema de atención (por ejemplo, baja resolutiveidad de los servicios, problemas de dotación de personal).

9. Etapa 1. Adquisición de datos

El primer paso es la recopilación de información a partir de diversos sistemas de información de salud disponibles, por ejemplo, del Registro Clínico Electrónico (RCE) Rayén. Esta información se podría complementar, con otras fuentes de datos como, encuestas de salud pública, información contenida en medios de comunicación (tanto escrito como no escrito), entre otras.

10. Etapa 2. Procesamiento de datos

En esta etapa se evalúa la disponibilidad e integridad de los datos y la presencia de datos perdidos (o ausentes). Además, se cautela la privacidad, anonimización y confidencialidad de la información. Esta etapa es crítica y de alto aliento, puesto que información de mala calidad en especial, en los registros de los textos libres, podría generar conclusiones incorrectas. Dentro de los problemas que se pueden presentar en los datos se encuentran: irregularidades numéricas, categorizaciones incorrectas, existencia de caracteres no numéricos en variables numéricas, ausencia de registros, entre otros. En base a estos posibles problemas, se deben definir criterios de exclusión de registros, así como considerar técnicas específicas de corrección de tales problemas.

11. Etapa 3. Exploración de datos

En esta etapa se obtienen los primeros resultados sobre el fenómeno de estudio. Sin embargo, no permite generar predicciones sobre el comportamiento futuro de los integrantes de la muestra. Con el fin de caracterizar la muestra en estudio se pueden comparar los datos con marcas de policonsultante v/s no policonsultante utilizando técnicas de resumen de datos y visualización (histogramas, boxplots, estadísticos de resumen, ANOVA).

12. Etapa 4. Análisis estadístico y modelamiento de datos

En esta etapa se explora el comportamiento predictivo diversos métodos tanto estadísticos como algorítmicos. Es aquí, donde ML entrega una batería de alternativas para evaluar supuestos con sustento estadístico, proponer métodos explicativos sólidos, y generar predicciones. En el caso de la detección de pacientes policonsultantes, dado la información de la variable de interés (número de asistencias a servicios de salud) se encuentra disponible en las bases de datos, se pueden utilizar algoritmos supervisados. Al finalizar la etapa, el grupo de investigación estará en posición de proponer una solución consistente para responder la (o las) pregunta de investigación propuesta.

13. Etapa 5. Validación y selección del modelo

Para realizar la validación y elección final del modelo, se deben considerar ciertas características asociadas a su desempeño entre ellas: su calidad predictiva y nivel de generalización. Para medir la calidad predictiva se pueden utilizar métricas provenientes de la matriz de confusión (tabla de contingencia), en conjunto con la curva ROC. Para la generalización del modelo una alternativa es aplicar del modelo a una nueva muestra integrada por individuos que el modelo, en su proceso de entrenamiento, no ha observado. Por ejemplo, otra muestra de pacientes que asisten a un servicio de urgencia.

Para definir umbrales de aprobación del modelo se debe tener presente la incidencia de consultantes en la población de estudio. No siempre es posible definir umbrales a priori y es aconsejable considerar la opinión de expertos en salud.

14. Etapa 6. Representación y reporte

Finalmente se operacionaliza la solución de manera funcional, por ejemplo, a través de un software . El empaquetamiento permitiría utilizar el modelo, o algoritmo, y en tiempo real, para predecir el comportamiento de un nuevo paciente categorizándolo como policonsultante o no. Además, hace posible evaluar escenarios simulados lo cual contribuye apoyando la toma de decisiones activas.

15. Conclusiones

La atención de salud genera un considerable volumen de datos, los cuales son difíciles de analizar con las técnicas estadísticas convencionales. En este sentido, las estrategias basadas en ciencia de datos y particularmente en ML resultan un recurso útil. Si bien estas aproximaciones han sido más utilizadas en el ámbito del diagnóstico clínico también representan una oportunidad para mejorar la gestión asistencial en las organizaciones de salud. Algunos ejemplos ilustrados en este trabajo son sus aplicaciones en la clasificación de pacientes y en la detección de usuarios policonsultantes. Actualmente, las bases de datos disponibles en nuestro país, como son, aquellas provenientes de los registros electrónicos e incluso de las imágenes almacenadas, ofrecen una importante oportunidad para generar conocimiento a partir de los datos existentes y retroalimentar a las organizaciones de salud contribuyendo a la toma de decisiones.

Referencias

- Wiemken TL, Kelley RR. Machine learning in epidemiology and health outcomes research. Annual Review of Public Health. 2020; 41 (1): 21-36.

- Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology*. 2017; 2 (4): 230-43.
- Andersen JKH, Pedersen JS, Laursen MS, Holtz K, Grauslund J, Savarimuthu TR, et al. Neural networks for automatic scoring of arthritis disease activity on ultrasound images. *RMD open*. 2019;5(1):e000891.
- Smits FT, Wittkamp KA, Schene AH, Bindels PJ, van Weert HC. Interventions on frequent attenders in primary care. *Scand J Prim Health Care*. 2008; 26 (2): 111-6.
- Rojas E, Muñoz-Gama J, Sepúlveda M, Capurro D. Process mining in healthcare: A literature review. *J Biomed Inform*. 2016; 61: 224-36.