

Diseño de un modelo para determinar la tecnología de predicción adecuada para diagnosticar la enfermedad renal crónica.

Yecid David Guzmán Díaz
Código estudiantil: 20071125242

Trabajo de Investigación presentado como requisito para optar el título de:
Magister en ingeniería de sistemas y computación

Tutor(es):

Phd, Juan Carlos Calabria Sarmiento

Phd Juan Manuel Rua Ascar

RESUMEN

Esta investigación aborda una de las tantas problemáticas del sector salud en Colombia, que se relaciona con el aumento de los casos de la enfermedad renal crónica y el alto costo para el sistema de salud. Además, destaca la falta de análisis de datos eficaz en el país, con escaso uso de técnicas de Machine learning o algoritmos de aprendizaje automático.

Para abordar estas deficiencias en el sistema de salud, primero se obtuvieron los datos de una institución prestadora de Salud (IPS) y se sometieron a un proceso exploratorio de

información. Durante la preparación de los datos, se llevaron a cabo procesos de selección, limpieza y transformación para garantizar la calidad de los datos para su posterior análisis mediante técnicas de minería de datos y algoritmos computacionales. Se utilizó la metodología CRISP-DM, que proporciona un enfoque completo y detallado para el proceso de minería de datos, y la distribución Anaconda de Python, una herramienta ampliamente utilizada para algoritmos de Machine learning.

El entrenamiento del modelo propuesto y diferentes modelos utilizaron un dataset supervisado dividido en un 70% para entrenamiento y el 30% para pruebas, además se usó un análisis de datos exploratorio para así conocer a detalle las características de las variables entregadas por la ips.

La métrica para determinar cuál es el más óptimo es la exactitud y la exhaustividad donde vemos que el modelo tiene un mejor rendimiento comparado diferente algoritmos de machine learning,

Dicha comparación define que el modelo propio es más exacto al momento de predecir la enfermedad renal crónica.

El análisis revela que las fechas impactan en todo el proceso, se encontraron diferencias significativas en los registros entre los distintos centros de atención de la IPS Previsalud, sobre todo dependiendo de los municipios y departamentos a los que están vinculados los usuarios objetos de dicha investigación.

También se concluyó que los procedimientos, laboratorios y signos vitales en el rango analizado representaron un porcentaje mayor en comparación con las consultas, hospitalizaciones y emergencias en el sector salud colombiano.

Además, la cantidad de consultas está relacionada con la calidad del rendimiento promedio, por lo que se convierte en un factor importante tanto en este estudio como en futuros análisis.

No se encontró una correlación significativa con las variables de grupo étnico y género.

Aunque los datos entregados no contaban la calidad deseada, se puede afirmar que el resultado de esta investigación es adecuado para los objetivos de negocio. Es importante destacar que esta investigación seleccionó con un doble propósito: establecer relaciones entre las variables y predecir la ocurrencia de la enfermedad renal crónica. Dado que el R2 es aproximadamente el 0.93, se considera un nivel alto de confiabilidad en comparación con la línea base.

Esta investigación puede implementarse de manera sencilla en cualquier proveedor de servicios en la nube. Para ilustrar este punto, se realizó una implementación básica de este modelo en Microsoft Azure Machine learning studio, que permite acceder al modelo a través de un servicio web. Utiliza las variables de entrada indicadas en el primer modelo de este trabajo y genera el pronóstico de la enfermedad renal crónica.

Palabras clave:

Enfermedad renal crónica, ERC, Machine learning, Sistema de salud colombiano, IPS, EPS Python, Azure Machine learning studio, modelos.

Key Words:

Chronic kidney disease, CKD, Machine learning, Health system colombiano, IPS, EPS, Python, Azure machine learning studio, Models

REFERENCIAS BIBLIOGRÁFICAS

- 1996, M. (s.f.). Obtenido de <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/DIJ/Decreto-2174-de-1996.pdf>
- Alba, A. O. (s.f.). Obtenido de <https://repository.javeriana.edu.co/handle/10554/51685>
- Anik, S., & Mittra, T. (2019). *Performance Measurements of Machine Learning Approaches for Prediction and Diagnosis of Chronic Kidney Disease (CKD)*. Obtenido de https://www.researchgate.net/publication/335850266_Performance_Measurements_of_Machine_Learning_Approaches_for_Prediction_and_Diagnosis_of_Chronic_Kidney_Disease_CKD
- Artal-Mittelmark, R. (2022). *Fisiología del embarazo*. Obtenido de <https://www.msmanuals.com/es/professional/ginecolog%C3%ADa-y-obstetricia/abordaje-de-la-mujer-embarazada-y-atenci%C3%B3n-prenatal/fisiolog%C3%ADa-del-embarazo>
- CAC. (s.f.). Obtenido de Cuenta de Alto Costo/ Situación de la Enfermedad Renal Crónica en Colombia, «MINISTERIO DE SALUD COLOMBIA,» 2019. [En línea]. Available: <https://acortar.link/B7iXG>. [Último acceso: 09 2020].
- Cardona, L. A. (2022). Obtenido de http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S2422-42002022000200125
- Care, F. M. (s.f.). Obtenido de Fresenius Medical Care Colombia S.A., «Enfermedad Renal: causas y prevención - Fresenius Medical Care,» 2019. [En línea]. Available: <https://acortar.link/Z2gKf>. [Último acceso: 09 2020]
- Colombia, C. d. (2012). Obtenido de <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=49981>
- Colombia, C. d. (2022). Obtenido de http://secretariasenado.gov.co/senado/basedoc/ley_2015_2020.html
- costo, C. d. (2023). Obtenido de <https://cuentadealtocosto.org/erc/dia-mundial-del-rinon-2023/>
- Cuenta de alto costo, C. (s.f.). *Qué es la cuenta de alto costo?* Obtenido de <https://cuentadealtocosto.org/quienes-somos/>
- Dario Soto, J. G. (2020). *Monitoreo de indicadores de valor a través de minería de datos, gestión de procesos de negocio y mejoramiento continuo con gestión del riesgo*. Obtenido de http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1692-33242020000200093

- Dibaba Adeba Debal, T. M. (2022). *Chronic kidney disease prediction using machine learning techniques*. Obtenido de <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00657-5>
- Europea, U. (2022). Obtenido de <https://universidadeuropea.com/blog/aprendizaje-supervisado-no-supervisado/>
- Evans, R. S. (2016). Obtenido de <https://pubmed.ncbi.nlm.nih.gov/27199197/>
- Fayyad, U. (1996). Obtenido de <https://dl.acm.org/doi/10.1145/240455.240464>
- Ferandez-Caballero, A., Manzano Arjona, M., Alonso González, E., & Tomé, S. (2006). *Una Perspectiva de la Inteligencia en su 50 aniversario*. España: La Roda.
- Francesco Paolo, A. V. (s.f.). Obtenido de <https://pubmed.ncbi.nlm.nih.gov/35543912/>
- Galán. (s.f.). Obtenido de <https://www.sngular.com/es/data-science-crisp-dm-metodologia/>
- Guersenzvaig, A. y. (s.f.). Obtenido de <https://dialnet.unirioja.es/servlet/autor?codigo=3193801>
- Isabel Ochoa, J. P. (2010). Obtenido de https://repositorio.uci.cu/jspui/bitstream/ident/TD_2516_09/1/TD_2516_09.pdf
- Khalid, T. A., Jalawi, A., Saud, A., & Solomon, A. (s.f.). Obtenido de <https://www.hindawi.com/journals/cin/2021/3941978/>
- Macros, D. (2022). Obtenido de <https://datosmacro.expansion.com/estado/gasto/salud/colombia#:~:text=El%20gasto%20p%C3%ABblico%20en%20sanidad,01%25%20del%20gasto%20p%C3%ABblico%20total.>
- Md. Ariful Islam, M. Z. (2023). *Chronic kidney disease prediction based on machine learning algorithms*. Obtenido de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9874070/>
- Michalski, B. y. (s.f.). Obtenido de <https://www.semanticscholar.org/paper/Machine-Learning-and-Data-Mining%3B-Methods-and-Michalski-Bratko/c21ce2fd906ef6b5c669d55458bd956155eda14f>
- Minsalud, C. (2013). *Guía de práctica clínica Hipertensión Arterial*. Obtenido de <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/CA/gpc-pacientes-hipertension-arterial-hta.pdf>
- Minsalud, C. (2015). *Clasificación Internacional de enfermedades CIE10*. Obtenido de <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/IA/SSA/cie10-cie11.pdf>
- Minsalud, C. (2015). *Preguntas frecuentes RIPS*. Obtenido de <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/OT/FAQ-RIPS.pdf>
- Minsalud, C. (2016). *Guía para consulta y elaboración de la ficha de regulación*. Obtenido de <https://www.minsalud.gov.co/Ministerio/Institucional/Procesos%20y%20procedimientos/GMTG03.pdf>

Minsalud, C. (2017). *Actualización Clasificación Única de Procedimientos en salud - CUPS*. Obtenido de <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VP/RBC/actualizacion-clasificacion-unica-procedimientos-salud-cups.pdf>

Minsalud, C. (s.f.). *Qué es la diabetes?* Obtenido de <https://www.minsalud.gov.co/salud/Paginas/Diabetes-minsalud.aspx>

Navarro, S. (2023). *Cómo programar inteligencia artificial? [5 lenguajes]*. Obtenido de <https://keepcoding.io/blog/como-programar-inteligencia-artificial/#:~:text=3%20%C2%BFC%C3%B3mo%20seguir%3F-,Algoritmos%2C%20clave%20de%20la%20IA,acci%C3%B3n%20o%20resolver%20un%20problema>

Pallarés Mestre, J. (2016). *La metodología cuantitativa aplicada al estudio de la reincidencia en menores infractores*. Obtenido de <https://dialnet.unirioja.es/servlet/tesis?codigo=150290>

Qué son las IPS? (2019). Obtenido de [https://sympliciasupport.zendesk.com/hc/es/articles/360001543071--Qu%C3%A9-son-las-IPS#:~:text=Son%20las%20Instituciones%20Prestadoras%20de,Plan%20Obligatorio%20de%20Salud%20\(POS\)](https://sympliciasupport.zendesk.com/hc/es/articles/360001543071--Qu%C3%A9-son-las-IPS#:~:text=Son%20las%20Instituciones%20Prestadoras%20de,Plan%20Obligatorio%20de%20Salud%20(POS))

R. A. Jeewantha, M., Halgamuge, A. M., & Ekici, G. (2017). *Classification performance analysis in medical science: Using kidney disease data*. Obtenido de <https://researchoutput.csu.edu.au/en/publications/classification-performance-analysis-in-medical-science-using-kidn>

Rajkumar Buyya, J. B. (2010). Obtenido de https://books.google.com.co/books/about/Cloud_Computing.html?id=S1NvRRd77rQC&redir_esc=y

republica, C. d. (1981). Obtenido de <https://funcionpublica.gov.co/eva/gestornormativo/norma.php?i=68760>

RIPS, P. f. (2019). Obtenido de <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/OT/FAQ-RIPS.pdf>

Rodriguez Barrios, R. H. (s.f.). Obtenido de <https://expeditiorepositorio.utadeo.edu.co/handle/20.500.12010/17638>

rodriguez, G. (s.f.). Obtenido de <https://repository.udistrital.edu.co/handle/11349/27876>

Rosselli, D. (2014). Obtenido de https://www.researchgate.net/publication/290530311_Respuesta_a_Analisis_de_costos_de_la_tamizacion_neonatal_universal_mediante_espectrometria_de_masas_en_tandem_para_errores_innatos_del_metabolismo_en_Colombia

salud, M. d. (19991). Obtenido de

https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/DIJ/Resoluci%C3%B3n_1995_de_1999.pdf#:~:text=RESOLUCION%20NUMERO%201995%20DE%201999%2028Julio%208%29%20Por,art%C3%ADculo%207%20del%20Decreto%201292%20de%201994%20y

Sharma, V., Stranieri, A., & Ugon, J. (2017). *An Agile Group Aware Process beyond CRISP-DM: A Hospital Data Mining Case Study*. Obtenido de <https://dl.acm.org/doi/abs/10.1145/3093241.3093273>

social, M. s. (2017). *Analisis De Situación De Salud (ASIS)*. Obtenido de <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/PSP/asis-nacional-2017.pdf>

Suárez Rozo, L. F., & Puerto García, S. (2017). *La crisis del sistema de salud colombiano: una aproximación desde la legitimidad y la regulación*. Obtenido de <https://revistas.javeriana.edu.co/index.php/gerepolsal/article/view/19879>

Surya, K., Kapeleshh, K., & Erik, D. (s.f.). *Machine Learning Prediction Models for Chronic Kidney Disease Using National Health Insurance Claim Data in Taiwan*. Obtenido de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8151834/>

Talamantes Alvarez, A. (2021). *Aprendizaje basado en ejemplos mediante el grafo*. Obtenido de https://cicese.repositorioinstitucional.mx/jspui/bitstream/1007/3603/1/Tesis_Ariana%20alamantes%20Alvarez_25%20agosto%202021.pdf

Uniandes. (s.f.). Obtenido de <https://uniandes.edu.co/es/noticias/salud-y-medicina/enfermedad-renal-cronica-erc-entre-las-de-alto-costo-en-colombia#:~:text=889.123%20casos%20de%20Enfermedad%20Renal,exceso%20de%20I%C3%ADquido%20del%20cuerpo>.

Vásquez Morales, G. R. (2019). *Clasificador con redes neuronales para e pronóstico de la enfermedad renal crónica en la población colombiana*. Obtenido de <https://reunir.unir.net/bitstream/handle/123456789/9446/V%C3%A1zquez%20Morales%20C%20Gabriel.pdf?sequence=1&isAllowed=y>

Vivas, H. (2015). Obtenido de <http://repositorio.unicauca.edu.co:8080/bitstream/handle/123456789/582/Optimizaci%C3%B3n%20en%20el%20entrenamiento%20del%20perceptr%C3%B3n%20multicapa.pdf?sequence=1&isAllowed=y>

Web, D. S. (2024). Obtenido de <https://www.datascience-pm.com/semma/>

William, A. M., & Floréz Valencia, L. (s.f.). *MODELOS DE APRENDIZAJE AUTOMÁTICO PARA LA PREDICCIÓN DE LA PROGRESIÓN DE LA ENFERMEDAD RENAL CRONICA*. Obtenido de https://repository.javeriana.edu.co/bitstream/handle/10554/63237/attachment_4_IEEE_Tesis.pdf?sequence=1&isAllowed=y

Zaforas, M. (2016). Obtenido de <https://www.paradigmadigital.com/dev/puede-aportar-big-data-al-mundo-la-medicina/>

Zúñiga-Pino, M.-Z. y-C. (s.f.). Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=5816932>