

Variables de Incidencia en la Mortalidad de Pacientes con Tuberculosis en el Departamento del Tolima

Wilson Andrés Gómez Gutiérrez^{1,*}

¹Especialización en Estadística Aplicada, Fundación Universitaria Los Libertadores

*Autor de correspondencia: wagomezg02@libertadores.edu.co.



Facultad de Ingeniería y
Ciencias Básicas



Resumen

La tuberculosis (TB) es la décimo tercera causa de muerte en el mundo, en Colombia el país ha adoptado la “Estrategia Mundial denominada Fin a la TB 2016-2035” a fin de mitigar el contagio y muertes por la enfermedad. En el análisis de la TB se han definido variables tales como la edad, sexo, las condiciones de salubridad y residencia del paciente. Variables que están asociadas al desarrollo más temprano de esta enfermedad, no obstante, no se ha confirmado que necesariamente sean estas quienes determinen una condición mortal en el paciente. Es por ello que a través de un modelo machine learning se determinan las características más importantes que están relacionadas con la evolución de la enfermedad TB y caracterizar los perfiles de pacientes con TB, de acuerdo a la información de las bases de datos de la plataforma de notificación de eventos en salud pública SIVIGILA y así poder estimar el porcentaje de mortalidad que puede llegar a tener un paciente de TB. Realizando la implementación se pudo mejorar el modelo base del modelo basado en reglas siendo el Quadratic Discriminant Analysis el mejor por sus métricas las cuales no son muy buenas pero tienen una tendencia de superar el modelo base.

Palabras clave: Tuberculosis, variables, salud pública, modelo de aprendizaje.

Recibido: 18 de junio de 2022
Aceptado: 2 de agosto de 2022
Publicado: 22 de noviembre de 2023

Como citar este artículo

Gómez-Gutiérrez, W. A., “Variables de Incidencia en la Mortalidad de Pacientes con Tuberculosis en el Departamento del Tolima”, *Revista Apuntes de Ciencia e Ingeniería*, 1, 1, nov, pag 5-10. 2023. doi: [10.37511/apuntesci.v1n1a1](https://doi.org/10.37511/apuntesci.v1n1a1)



Copyright: ©2023 por los autores. Este artículo es de acceso abierto distribuido bajo los términos y condiciones de Creative Commons Licencia de atribución (CC BY NC SA) (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

1. Introducción

La tuberculosis, actualmente es considerada una de las principales enfermedades causante de la mortalidad de millones de personas.

Según el Informe Mundial de Tuberculosis 2019 de la Organización Mundial de la Salud (OMS), alrededor de 1700 millones de personas infectadas con *Mycobacterium tuberculosis* desarrollarán tuberculosis durante su vida [1, 4, 5, 6].

Se han estudiado variables relacionadas con las condiciones del ambiente en el que viven los pacientes con TB que inciden en el desarrollo de la enfermedad y posiblemente incrementan las posibilidades de muerte en los pacientes, según la investigación realizada por el Dr Walter H. Curioso y Maria J, Brunette, publicada en la revista Scielo Perú [2], además se estudia la variable correspondiente a los ingresos económicos del paciente que le facilitan el acceso al servicio de salud, compra de medicamentos y el tratamiento pronto a la enfermedad.

Los pacientes de TB deben de adoptar medidas preventivas frente a su autocuidado e higiénicas para prevenir la propagación y aceleración de la enfermedad. En el desarrollo de tesis enfocada en “*Conocimientos de las Medidas Preventivas y Actitudes en el Autocuidado de pacientes con Tuberculosis Pulmonar en el Centro de Salud Los Libertadores*” [3]. Se concluye que la variable de ingreso en los pacientes con TB es de 83,75 % con relación a (34) muestras a las que se les aplicó el instrumento, de las cuales, tienen conocimientos sobre los cuidados en el hogar, pero si no cuentan con una vivienda adecuada y medios económicos necesarios no podrán cuidar su salud.

En el desarrollo del presente trabajo de investigación titulado “*Variables de incidencia en la mortalidad de pacientes con tuberculosis en el Tolima*” se busca identificar los factores individuales y del entorno que pueden pronosticar la mortalidad de un paciente con TB con la ayuda de modelos de aprendizaje automático supervisado. Para tal fin, se realizará el análisis de estudio con los datos privados de las bases cerradas del año 2019, 2020 y 2021, proporcionados por el Sistema de Salud Pública (SIVIGILA).

La investigación resultará ser un instrumento de bastante importancia en el sector de la salud, porque permitirá caracterizar al paciente de acuerdo con las variables y esto servirá para establecer estrategias de detección e intervención temprana según el perfil del paciente.

2. Metodología

El presente estudio se basó en los datos privados de las bases cerradas del año 2019, 2020 y 2021, proporcionados por SIVIGILA el Sistema de Vigilancia en Salud Pública. Solicitados al coordinador de SIVIGILA y estadísticas vitales del Tolima, disponible en [7].

SIVIGILA es la plataforma estatal de notificaciones de eventos en salud pública que se encarga de recopilar toda la información de eventos en salud pública de las entidades de salud prestadoras del servicio; siendo estas las encargadas de alimentar todas bases y así tener consolidado los casos que se presenten para realizar posteriormente el control epidemiológico llegase al caso y poder establecer planes de contingencia y demás protocolos que se requieran para el manejo de los eventos.

Estas bases son descargadas de manera anual el primer trimestre del año siguiente a la construcción de la misma y correspondiente a los 3 años mencionados se contó con 1.378 registros de personas con TB (ver descripción descripción en la Tabla 1).

Las fases para llevar a cabo el presente estudio fueron:

- Limpieza de la base de datos y selección de los registros relevantes
- Realización de análisis descriptivos
- Identificación de posibles modelos y la respectiva evaluación de su desempeño.

En el análisis descriptivo de la base de tuberculosis (TB) encontramos, $n = 1372$ con 24 variables. Las variables están distribuidas en variables dicotómicas y categóricas.



Al realizar un análisis descriptivo de los datos encontramos que la variable edad no cuenta con una normalidad; por ello se realiza una transformación aplicando logaritmo base 10 creando una nueva variable edad_log10 en la tabla df2 creada a través de python.

La variable estrato presenta una ausencia del 8% de sus datos respecto al total de registros con pacientes de TB, pero según análisis no es representativo para continuar con esa variable que explica un comportamiento del estrato respecto a la mortalidad por TB que es la variable objetivo.

Para este estudio las variables categóricas cuentan con un valor cuando 1 es afirmativo al significado de la variable. A continuación, en las Figura 1 y 2 se exponen las variables, sus significados y transformaciones realizadas.

Variable	Descripción y unidades
año	año de notificación del paciente
edad	edad de la persona (Se aplica Log 10)
uni_med_	Unidad de medida de la edad
sexo_	sexo de la persona
area_	Área de residencia. 1: Cabecera Municipal 2: Centro poblado 3: Rural disperso
estrato_	Estrato de la persona
pac_hos_	paciente hospitalizado. 1: SÍ 0: No
con_fin_	condición final del paciente 0: vivo 1: Muerto
con_tuber	condición de la tuberculosis 1: Sensible 0: Resistente
tip_tub	Tipo de tuberculosis 1: Pulmonar 0: Extrapulmonar
clas_ant	Clasificación anterior 1: Nuevo 0: Previamente tratado
vih_confirmado	Confirmación VIH 1: SI 0:NO
peso_act	Peso Actual persona en Kg
talla_act	Talla actual persona en Metros
imc	Índice de masa corporal persona
diabetes	diabetes 1: SI 0:NO
silicosis	Silicosis 1: SI 0: NO
enfe_renal	Enfermedad renal 1: SI 0: NO
epoc	Epoc 1: SI 0: NO
enfe_hepat	Enfermedad hepática 1: SI 0: NO
cancer	Cáncer 1: SI 0: NO
artritis	Artritis 1: SI 0: NO
desnutrición	desnutrición 1: SI 0: NO
nmun_proce	Municipio de Procedencia

Tabla 1: Transformación de algunas de las variables previa al análisis.

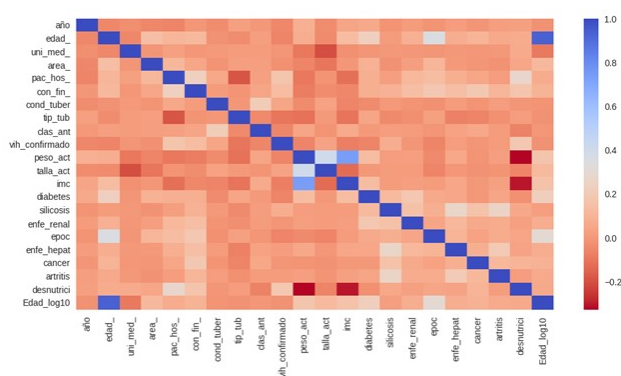


Figura 1: Gráfica de correlación de variables por TB.



Se puede evidenciar la relación o correlación que tienen algunas variables respecto a la variable objetivo `can_fin_` para determinar seguidamente los posibles modelos a implementar. el Ranking de variables que resultó del análisis fueron:

1. Silicosis = `silicosis_`
2. Enfermedad Hepática = `enfe_hepat`
3. Cancer = `cancer_`
4. Edad = `edad_`
5. Enfermedad renal = `enfe_renal`
6. Epoc = `epoc_`
7. Artritis = `artritis_`
8. Desnutrición = `desnutricion_`

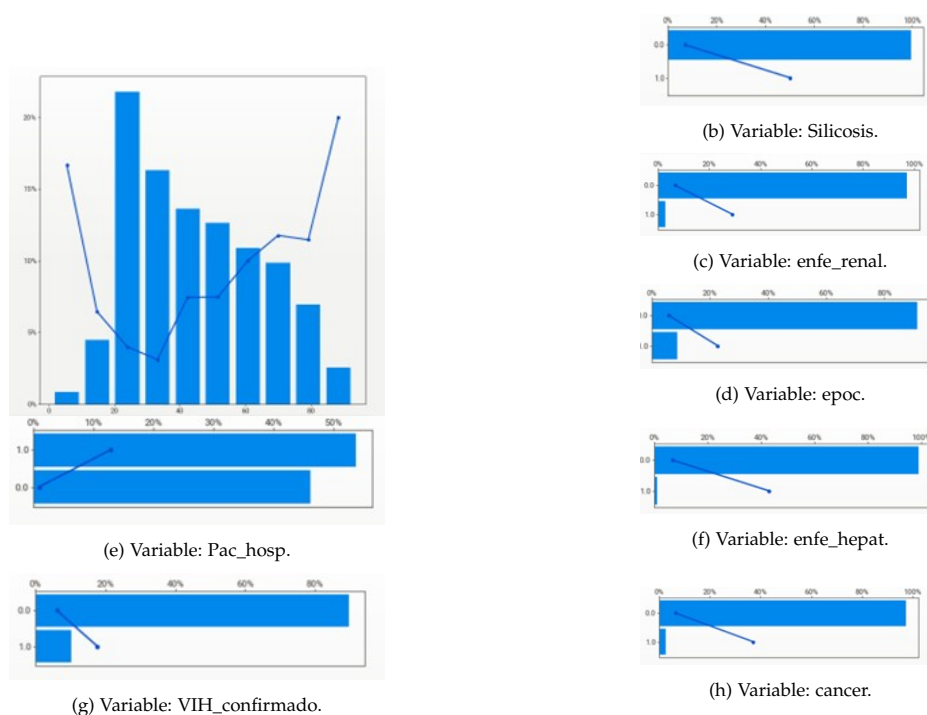


Figura 2: Una matriz de imágenes.

Al observar las variables estas no presentan una gran variabilidad ya que existen muy pocos registros donde la persona fallece a causa de la variable. Se estima conveniente aplicar el balanceo de clases. A través de Python y la librería **PyCaret** se realizan dos modelos el cual muestra que a mayor número de variables se mejora poco a poco el modelo ya que se tienen pocos registros de fallecimientos.

3. Resultados

3.1. Creación inicial de modelo basado en reglas:

Se realizó la separación de la tabla en entrenamiento y testeo donde obtenemos los `X_train`, `y_train`, `X_test`, `y_test`. Seguidamente se crea un modelo de ajuste y predicción con una sola variable para obtener modelo de referencia y partir a mejorar y comprar con el modelo machine learning que se piensa realizar;



donde obtenemos un accuracy = 94.92 pero con ella una precisión = 0.00 (ver Figuras 3 y 4).

A continuación, las Tablas 2, 3 y 4 presentan los resultados del modelo machine learning el cual arrojó mejores resultados en comparación con el modelo basado en reglas.

	precision	recall	f1-score	support
0	0.95	1.00	0.97	262
1	0.00	0.00	0.00	14

Tabla 2: Resultados iniciales modelo basado en reglas.

Se comparan los modelos y obtenemos la siguiente tabla de modelos:

Model	Accuracy	AUC	Recall	Prec.	F1
Quadratic Discriminant Analysis	0.8897	0.6535	0.3095	0.3071	0.3385
Linear Discriminant Analysis	0.9001	0.6525	0.1048	0.3583	0.1433
Ada Boost Classifier	0.9157	0.6523	0.0429	0.35	0.0722
Logistic Regression	0.9157	0.6488	0.0286	0.3	0.05
Naive Bayes	0.8884	0.6478	0.3095	0.3018	0.3357

Tabla 3: Comparación de modelo final.

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Quadratic Discriminant Analysis	0.9305	0.7947	0.45	0.4286	0.439	0.402	0.4021

Tabla 4: Predicción del modelo.

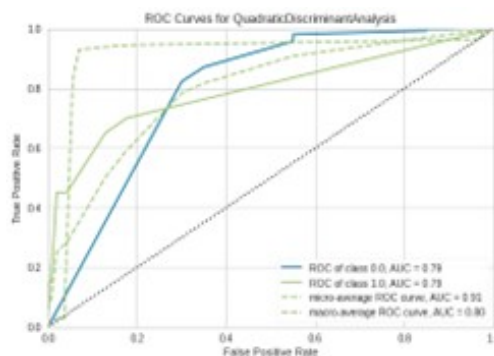


Figura 3: Curva ROC.

4. Discusión de resultados

De acuerdo con la metodología planteada se mejora el modelo basado en reglas; ya que este no mostraba ninguna precisión al respecto para evaluar las condiciones de los pacientes por TB, este mostraba que de 27 personas fallecidas solo podía predecir a 1.

Se crea un modelo inicial el cual tiene una pequeña mejora ilustrada en una matriz de confusión donde de 20 casos el modelo alcanza a clasificar correctamente 9 de ellos.

Estos resultados tienen un gran problema y es el sesgo que tienen los datos ya que no se contaba con un número significativo de muertes por TB lo que sesgaba como tal la base. Se aplicó en un momento dado el balanceo de clases, pero este empeoró como tal el modelo haciendo que su predicción respecto a las muertes y a las personas que vivieron fuera de 0.



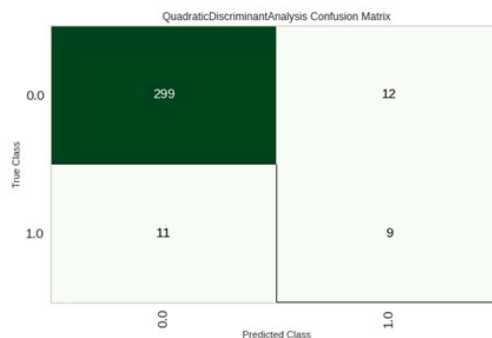


Figura 4: Matriz de confusión.

5. Conclusiones

Para este ejercicio este instrumento se puede tomar en cuenta para un pequeño análisis inicial de prioridad a un paciente diagnosticado con TB, ya que arroja resultados de certeza de más del 50%; pero aclarando que no debe ser decisivo en la priorización de pacientes con TB. Este instrumento debe ser evaluado nuevamente para poder establecer a través de los reportes de las instituciones de salud, que variables pueden ser más incisivas en la predicción de mortalidad por TB; esto con el fin de poder actuar de manera pronta y poder clasificar la prioridad de paciente creando una oportunidad de que podamos disminuir el riesgo de mortalidad y salvar más vidas con una atención temprana y priorizada.

Referencias

- [1] Becker, A., Blüthgen, C., van, V., Sekaggya-Wiltshire, C., Castelnuovo, B., Kambugu, A., Fehr, J., and Frauenfelder, T. (2018). Detection of tuberculosis patterns in digital photographs of chest x-ray images using deep learning: Feasibility study. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease*, 22:328–335, DOI: [10.5588/ijtld.17.0520](https://doi.org/10.5588/ijtld.17.0520).
- [2] Curioso, W. H. and Brunette, M. J. (2020). Inteligencia artificial e innovación para optimizar el proceso de diagnóstico de la tuberculosis. *Revista Peruana de Medicina Experimental y Salud Pública*, 37(3):554–8, DOI: [10.17843/rpmesp.2020.373.5585](https://doi.org/10.17843/rpmesp.2020.373.5585), <https://doi.org/10.17843/rpmesp.2020.373.5585>.
- [3] Larico Cruz, C. M. (2009). Conocimientos de las medidas preventivas y actitudes en el autocuidado de pacientes con tuberculosis pulmonar en el centro de salud los libertadores en noviembre–diciembre 2008.
- [4] Melendez, J., Sánchez, C. I., Philipsen, R. H. H. M., Maduskar, P., Dawson, R., Theron, G., Dheda, K., and van Ginneken, B. (2016a). An automated tuberculosis screening strategy combining x-ray-based computer-aided detection and clinical information. *Scientific reports*, 6:25265, ISSN: 2045-2322, DOI: [10.1038/srep25265](https://doi.org/10.1038/srep25265).
- [5] Melendez, J., Sánchez, C. I., Philipsen, R. H. H. M., Maduskar, P., Dawson, R., Theron, G., Dheda, K., and van Ginneken, B. (2016b). An automated tuberculosis screening strategy combining x-ray-based computer-aided detection and clinical information. *Scientific reports*, 6:25265, ISSN: 2045-2322, DOI: [10.1038/srep25265](https://doi.org/10.1038/srep25265), <https://europepmc.org/articles/PMC4850474>.
- [6] Organization, W. H. (2019). *Global tuberculosis report 2019*. World Health Organization.
- [7] SIVIGILA (2022). Data: Sistema de vigilancia en salud pública. Disponible en: <https://colab.research.google.com/drive/> Accedido: 16 de mayo de 2022".

