



Unidad Pedagógica
San Ignacio Guazú

CARRERA DE INGENIERÍA INFORMÁTICA

TRABAJO FINAL DE GRADO

PREDICCIÓN DE LA DISPERSIÓN GEOGRÁFICA DE ENFERMEDADES
RESPIRATORIAS CRÓNICAS MEDIANTE CIENCIA DE DATOS

2025



Unidad Pedagógica
San Ignacio Guazú

CARRERA DE INGENIERÍA INFORMÁTICA

TRABAJO FINAL DE GRADO

PREDICCIÓN DE LA DISPERSIÓN GEOGRÁFICA DE ENFERMEDADES
RESPIRATORIAS CRÓNICAS MEDIANTE CIENCIA DE DATOS

GUILLERMO ZACARIA ARGÜELLO CORDOBE

Tutores

Ing. Santiago Raúl Caballero Semenaka

Ing. Aura Rocío Villalba Alderete

2025

Resumen

Este trabajo final de grado tuvo como objetivo facilitar la predicción de la dispersión geográfica de enfermedades respiratorias crónicas mediante la aplicación de ciencia de datos, buscando identificar perfiles de riesgo, localizar zonas geográficas de alta concentración y predecir la demanda futura de consultas en las distintas unidades de salud familiar. Se aplicó un enfoque mixto, diseño exploratorio e investigación de tipo aplicada, contando con la colaboración del departamento de bioestadística e informática del Hospital Integrado de Ayolas quienes proporcionaron datos reales de consultas correspondiente al periodo 2020 a 2025, procesados con Python y librerías como Pandas y Scikit-learn. Se implementó un modelo de clustering con K-Means para la segmentación de pacientes y se realizó una evaluación comparativa de cinco modelos de regresión para el pronóstico de series de tiempo. Todos los hallazgos se integran en una aplicación de escritorio funcional desarrollada en Tkinter, conectada a una base de datos MySQL. Como resultado, se obtuvieron tres perfiles de riesgo claros validados con un excelente Coeficiente de Silueta de 0.92. En la predicción de consultas, los modelos lineales Lasso y ElasticNet demostraron un rendimiento superior, alcanzando un R^2 de 0.97 en el pronóstico general y un error promedio (MAE) de solo 2 a 5 consultas. La herramienta desarrollada provee al personal de salud mapas interactivos, perfiles de riesgo y pronósticos fiables, permitiendo una transición de una gestión reactiva a una proactiva. Como trabajo futuro, se propone la automatización de la base de datos y la migración de la solución a una plataforma web para mejor accesibilidad por parte de los funcionarios.

Palabras clave: Ciencia de Datos, Salud Pública, Enfermedades Respiratorias Crónicas, Modelado Predictivo, Clustering.

Abstract

This final degree project aimed to facilitate the understanding of chronic respiratory diseases through the application of data science, seeking to identify risk profiles, locate geographic areas of high concentration, and predict the future demand for consultations. A mixed-method approach, exploratory design, and applied research were used, with the collaboration of the biostatistics and informatics department of the Ayolas Integrated Hospital, which provided real consultation data from the 2020 to 2025 period, processed using Python and libraries such as Pandas and Scikit-learn. A K-Means clustering model was implemented for patient segmentation, and a comparative evaluation of five regression models was performed for time series forecasting. All findings are integrated into a functional desktop application developed in Tkinter, connected to a MySQL database. As a result, three clear risk profiles were obtained, validated with an excellent Silhouette Coefficient of 0.92. In consultation forecasting, the linear models Lasso and ElasticNet demonstrated superior performance, achieving an R^2 of 0.97 in the general forecast and an average error (MAE) of only 2 to 5 consultations. The developed tool provides health staff with interactive maps, risk profiles, and reliable forecasts, enabling a transition from reactive to proactive management. As future work, the automation of the database and the migration of the solution to a web platform are proposed for better accessibility by staff.

Keywords: Data Science, Public Health, Chronic Respiratory Diseases, Predictive Modeling, Clustering.

1. Introducción

Las enfermedades respiratorias crónicas constituyen uno de los principales problemas de salud pública no solo en la ciudad de Ayolas sino a nivel mundial, debido a su alta prevalencia. (Organización Mundial de la Salud (OMS), 2019) Entre las más frecuentes se encuentran el Asma, la enfermedad pulmonar obstructiva crónica (EPOC), la bronquitis crónica y otras patologías que, aunque no siempre se manifiestan con la misma intensidad, suelen ser progresivas y requieren un control continuo. En comunidades rurales y urbanas, estas enfermedades están asociadas a factores de riesgo como el tabaquismo, la exposición al humo de leña, la contaminación ambiental y las condiciones socioeconómicas.

El diagnóstico y seguimiento de estas afecciones suele depender de consultas médicas presenciales, estudios clínicos, radiológicos y pruebas de función respiratoria que, si bien resultan efectivos, no siempre están disponibles para toda la población. En consecuencia, el acceso tardío a la detección y tratamiento puede derivar en complicaciones que incrementan la mortalidad y los gastos hospitalarios. Frente a esta problemática, surgió la necesidad de implementar soluciones tecnológicas que permitieran analizar grandes volúmenes de información clínica y epidemiológica de manera eficiente, ofreciendo una visión más clara sobre el comportamiento y la evolución de estas enfermedades en diferentes sectores de la población ayolense.

El proyecto contempló la implementación de un sistema que recopiló datos provenientes del Hospital Integrado de la ciudad de Ayolas, específicamente de la unidad de Bioestadística e Informática. Dichos datos incluyeron variables clínicas y demográficas como edad, sexo, lugar de residencia, aseguramiento, motivos de consulta y diagnósticos codificados en CIE-10, lo que posibilitó un análisis integral de la población atendida.

Asimismo, la investigación abarcó fases fundamentales de la ciencia de datos que incluyeron la limpieza y organización de los datos, así como el desarrollo de modelos descriptivos y predictivos. Para la visualización y análisis de los resultados, se implementó un sistema en Python y Tkinter, que permitió la representación de mapas de calor, la evolución temporal de casos y la segmentación por perfiles de riesgo. De esta manera, se buscó no solo analizar las tendencias de las enfermedades respiratorias crónicas en la región, sino también anticipar posibles escenarios futuros para facilitar la planificación de recursos médicos y campañas preventivas.

En cuanto al valor de esta investigación, se consideró altamente relevante, ya que permitió disponer de un sistema accesible y confiable para apoyar la toma de decisiones clínicas y de

gestión sanitaria. El proyecto contribuyó significativamente a optimizar la eficiencia de los servicios de salud locales, facilitando el diagnóstico oportuno y la vigilancia de las enfermedades respiratorias crónicas.

1.1. Planteamiento del problema

Las enfermedades respiratorias crónicas, como el Asma y la Enfermedad Pulmonar Obstructiva Crónica (EPOC), representan un importante problema de salud pública debido a su alta prevalencia y al impacto que generan en la calidad de vida de los pacientes. La gravedad de este problema es global: la EPOC es la cuarta causa de muerte en el mundo y en 2021 ocasionó 3.5 millones de defunciones, lo que representa aproximadamente el 5% de todas las defunciones a nivel mundial. (OMS)

En la ciudad de Ayolas, el Hospital Integrado dispone de registros de pacientes que podrían resultar de gran valor para el análisis de estas enfermedades. Sin embargo, la información se encuentra fragmentada en diferentes archivos, sin un sistema unificado que permita integrar y aprovechar adecuadamente los datos. Esta situación dificulta identificar patrones de comportamiento, conocer los grupos de mayor riesgo y ubicar las zonas con mayor concentración de casos. En consecuencia, se limita la posibilidad de planificar recursos de manera eficiente, mejorar la atención a los pacientes y diseñar campañas preventivas.

Ante este contexto, surge la necesidad de implementar herramientas de Ciencia de Datos y que permitan transformar los registros clínicos, demográficos y geográficos en información clara y útil. A través de gráficos, mapas y tableros interactivos, poder visualizar tendencias, comprender mejor la distribución de los casos y aplicar modelos predictivos que anticipen posibles incrementos en determinadas zonas o poblaciones.

El problema central radica en que los datos actualmente disponibles no están integrados ni se explotan de manera sistemática, lo que limita la generación de conocimiento aplicable para la toma de decisiones en salud pública. Este proyecto plantea una alternativa de solución mediante la construcción de un sistema que organice, depure y analice la información, ofreciendo resultados accesibles para profesionales de la salud y gestores. De esta manera, se contribuirá a optimizar la comprensión de las enfermedades respiratorias crónicas en la población de Ayolas y a fortalecer las estrategias de prevención y atención médica.

1.2. Justificación

Este proyecto responde a la necesidad de contar con herramientas que permitan analizar de manera clara y ordenada las enfermedades respiratorias crónicas, como el Asma y la EPOC, que afectan significativamente la calidad de vida de muchas personas. La falta de integración de los datos en los hospitales hace que sea difícil identificar patrones, zonas de riesgo y poblaciones vulnerables, lo cual limita la prevención y la atención médica oportuna.

La aplicación de Ciencia de Datos ofrece una alternativa moderna y eficaz para transformar datos dispersos en información comprensible y útil. Mediante el uso de estas tecnologías, se podrán construir gráficos, mapas y paneles interactivos que faciliten a los profesionales de la salud la interpretación de los registros médicos y la planificación de recursos. Además, el análisis predictivo permitirá anticipar el aumento de casos en determinados barrios o grupos de pacientes, mejorando así la respuesta ante situaciones críticas.

El principal aporte de esta investigación es desarrollar una solución accesible y práctica que apoye tanto al personal de salud como a las autoridades sanitarias en la toma de decisiones. Este proyecto no solo ayudará a organizar y visualizar mejor la información disponible en el Hospital Integrado de Ayolas, sino que también abrirá el camino para futuras investigaciones y aplicaciones en otros centros de salud con recursos limitados.

Esta investigación busca ser una herramienta que combine innovación tecnológica con un impacto social directo, mejorando la prevención, el diagnóstico y la calidad de vida de los pacientes con enfermedades respiratorias crónicas.

1.3. Alcance y limitaciones

1.3.1. Alcance

Este trabajo tuvo como finalidad desarrollar e implementar un sistema basado en herramientas de ciencia de datos para el análisis de enfermedades respiratorias crónicas, como el asma y la EPOC, utilizando datos clínicos, demográficos y geoespaciales reales. El trabajo integró procesos de recolección, limpieza, análisis, visualización y predicción de datos relevantes para la toma de decisiones en salud pública.

La solución se materializó en un Sistema de Apoyo a la Decisión (DSS) desarrollado en Python, utilizando la librería Tkinter para la interfaz gráfica y MySQL para la gestión persistente de los datos. Este sistema incluyó tableros de mando (dashboards) para la presentación de los principales indicadores clave de rendimiento para la gestión sanitaria y la vigilancia epidemiológica. Asimismo, contó con módulos de análisis geoespacial y visualización interactiva, incluyendo mapas de calor generados para identificar los focos críticos de

enfermedades por barrio. Se incorporó la segmentación de riesgo mediante la implementación de un modelo de clustering (K-Means) para agrupar pacientes en perfiles de riesgo homogéneos. Finalmente, el sistema albergó un modelo predictivo avanzado de series de tiempo que permitió anticipar el comportamiento futuro de estas enfermedades en distintas zonas o grupos poblacionales.

1.3.2. Límites

El proyecto se centró únicamente en enfermedades respiratorias crónicas y no incluyó el análisis de enfermedades respiratorias agudas (como influenza o COVID-19). No se integraron módulos avanzados de diagnóstico clínico automatizado ni procesamiento de imágenes médicas. Debido a que la conectividad con los sistemas hospitalarios dependió de las condiciones técnicas de la institución, la carga de datos se realizó de forma manual o semiautomática. Además, no se contempló la integración de datos en tiempo real ni se establecieron conexiones activas con sistemas hospitalarios o expedientes médicos electrónicos.

1.4. Objetivos

1.4.1. Objetivo general

Predecir la dispersión geográfica de enfermedades respiratorias crónicas mediante el uso de herramientas de ciencia de datos.

1.4.2. Objetivos específicos

- 1) Entender el impacto de las enfermedades respiratorias crónicas según variables demográficas y geográficas en la población de estudio.
- 2) Identificar zonas geográficas con alta concentración de casos, mediante mapas de calor de casos por barrios, para dirigir campañas de prevención.
- 3) Detectar perfiles de riesgo agrupando pacientes con características similar

1.5. Contenido del Trabajo Final de Grado

- **Introducción:** Se presentó el tema de investigación, la importancia de estudiar las enfermedades respiratorias crónicas y el impacto que tienen en la salud pública. Asimismo, se expusieron el objetivo general y los objetivos específicos, junto con la motivación de aplicar Ciencia de Datos y Aprendizaje Automático para transformar los registros médicos en información útil para la toma de decisiones.
- **Marco Teórico:** Describe el estado del arte en el uso de ciencia de datos en el área de salud, la definición de enfermedades respiratorias crónicas (asma, EPOC, bronquitis, etc.) y los conceptos fundamentales relacionados con el análisis descriptivo, predictivo, exploratorio, tableros de mando y reportes interactivos de KPIs.
- **Marco Metodológico:** Se detallaron las técnicas, herramientas y procedimientos aplicados en el proyecto. Incluye el proceso de recolección de datos hospitalarios, la preparación y depuración de los mismos, la construcción de la base de datos en MySQL, el análisis con Python y la visualización especializada mediante la interfaz Tkinter. Además, se presentaron los resultados de los análisis exploratorios y de los modelos predictivos aplicados.
- **Conclusiones:** Se destacaron los hallazgos más importantes del estudio, como la identificación de zonas críticas, perfiles de riesgo y patrones demográficos de los pacientes. Se resaltó el aporte de la Ciencia de Datos para mejorar la gestión de información en salud y el apoyo del sistema a la toma de decisiones médicas y administrativas.
- **Trabajos Futuros:** Se plantearon líneas de investigación y mejoras, como la ampliación del sistema a otros centros hospitalarios, la integración de datos en tiempo real y la incorporación de técnicas avanzadas de Machine Learning.