



Atención hospitalaria de pacientes con COVID-19

¿Se pueden agrupar los pacientes según sus características de ingreso?

Proyecto Final - Análisis no supervisado



Bienvenidos al Proyecto de Agrupación de Pacientes COVID-19

Exploraremos los datos de pacientes que han sido hospitalizados por COVID-19 en cuatro clínicas en diferentes ciudades de Colombia. Estos datos incluyen detalles demográficos, síntomas al momento del ingreso, variables de examen físico, clasificaciones clínicas y resultados (ingreso a la unidad de cuidados intensivos-UCI y mortalidad).

Anteriormente, nuestro equipo utilizó estos datos para construir modelos supervisados con el objetivo de predecir ingresos a la UCI. En este proyecto, adoptamos un enfoque diferente al aplicar técnicas de aprendizaje no supervisado.

Objetivos

Nuestra meta es reducir la dimensionalidad y descubrir grupos distintos de pacientes basados en sus características al momento de la hospitalización. Esto podría ser valioso para ajustar los tratamientos iniciales según las características de cada grupo.

Potenciales Partes Interesadas

Varios actores podrían encontrar este análisis beneficioso, incluyendo:

- Clínicas y hospitales (públicos y privados)
- Administradores de planes de beneficios de salud
- Consortios de prestadores de servicios de salud
- Profesionales médicos, enfermeras y personal de salud
- Pacientes
- Comunidad en general



Muchos de estos actores podrían ser clientes directos interesados en implementar procesamiento y análisis similares en sus instituciones.

En este proyecto, la fuente de datos consiste en cuatro clínicas privadas en Colombia. La información se obtuvo con la aprobación institucional y la autorización de un comité de ética en investigación.

Descripción de archivos

Este proyecto incluye archivos ejecutables, archivos de texto y directorios de la siguiente manera:

Archivos Ejecutables:

- **modelo.ipynb** - Incluye todas las funciones necesarias para las operaciones de agrupación. Utiliza las funciones definidas para generar el archivo `evaluation.txt` después de la ejecución.

Archivos de Salida:

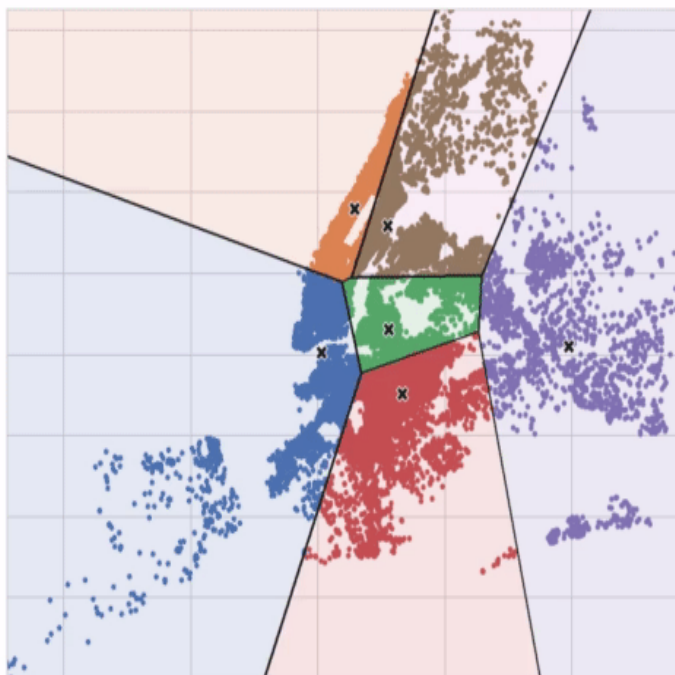
- **evaluation.txt** - Contiene una tabla de resultados por paciente y su respectiva agrupación.

Directorios de Origen:

- **directorio de entrenamiento** - Contiene el dataset de pacientes "BD_Morbimortalidad-COVID-19-2.xlsx"

Metodología

La agrupación (clustering) es una técnica de aprendizaje no supervisado en la que los datos se dividen en clústeres basados en similitudes. El objetivo principal es capturar patrones subyacentes en los datos sin necesidad de un labeling previo. En el contexto de nuestro dataset, el clustering puede ayudar a identificar grupos de pacientes con síntomas similares, lo que puede ser valioso para el análisis y la toma de decisiones en los centros de atención médica.



Para este proyecto, consideraremos inicialmente los siguiente algoritmos de clustering para elaborar el modelo:

- **K-Means**
- **K-Medoides**
- **Clustering Jerárquico**
- **DBSCAN**

La elección del algoritmo final dependerá de la naturaleza de los datos y de la evaluación de los resultados para determinar cuál es más adecuado .

Instrucciones para ejecución del model

El orden de ejecución de los archivos del programa es el siguiente:

1) model.ipynb

Se ejecuta para efectuar la creación del archivo evaluation.txt. Al principio de este archivo, se ha importado "BD_Morbimortalidad-COVID-19-2.xlsx" para que las funciones definidas en él puedan ser utilizadas.

2) evaluation.txt

Finalmente, se debe ejecutar el archivo texto para la comprobación de resultados.

Bibliografía

- **Dong E, Du H, Gardner L.** An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis.* 2020;20(5):533-4.
- **Collaborators C-CI.** Estimating global, regional, and national daily and cumulative infections with SARS-CoV-2 through Nov 14, 2021: a statistical analysis. *Lancet.* 2022;399(10344):2351-80.
- **Lin C-T, Bookman K, Sieja A, Markley K, Altman RL, Sippel J, et al.** Clinical informatics accelerates health system adaptation to the COVID-19 pandemic: examples from Colorado. *Journal of the American Medical Informatics Association.* 2020;27(12):1955-63.
- **Merette K, Awad M, Hamid R.** Building resilient hospitals in the Eastern Mediterranean Region: lessons from the COVID-19 pandemic. *BMJ Global Health.* 2022;7(Suppl 3):e008754.
- **Staffolani S, Iencinella V, Cimatti M, Tavio M.** Long COVID-19 syndrome as a fourth phase of SARS-CoV-2 infection. *Infez Med.* 2022;30(1):22-9.
- **Saavedra Trujillo CH.** ADENDO: ACTUALIZACIÓN 27/06/2020. Consenso colombiano de atención, diagnóstico y manejo de la infección por SARS-CoV-2/COVID-19 en establecimientos de atención de la salud: Recomendaciones basadas en consenso de expertos e informadas en la evidencia ACIN-IETS. SEGUNDA EDICIÓN. *Infectio.* 2020;24(3).
- **Mirfazeli FS, Sarabi-Jamab A, Jahanbakhshi A, Kordi A, Javadnia P, Shariat SV, et al.** Neuropsychiatric manifestations of COVID-19 can be clustered in three distinct symptom categories. *Scientific Reports.* 2020;10(1):20957.
- **Otake S, Chubachi S, Namkoong H, Nakagawara K, Tanaka H, Lee H, et al.** Clinical clustering with prognostic implications in Japanese COVID-19 patients: report from Japan COVID-19 Task Force, a nation-wide consortium to investigate COVID-19 host genetics. *BMC Infectious Diseases.* 2022;22(1):735.
- **Fernández-de-las-Peñas C, Martín-Guerrero JD, Florencio LL, Navarro-Pardo E, Rodríguez-Jiménez J, Torres-Macho J, et al.** Clustering analysis reveals different profiles associating long-term post-COVID symptoms, COVID-19 symptoms at hospital admission and previous medical comorbidities in previously hospitalized COVID-19 survivors. *Infection.* 2023;51(1):61-9.
- **San-Cristobal R, Martín-Hernández R, Ramos-Lopez O, Martinez-Urbistondo D, Micó V, Colmenarejo G, et al.** Longwise Cluster Analysis for the Prediction of COVID-19 Severity within 72 h of Admission: COVID-DATA-SAVE-LIFES Cohort. *Journal of Clinical Medicine.* 2022;11(12):3327.
- **Hu F, Huang M, Sun J, Zhang X, Liu J.** An analysis model of diagnosis and treatment for COVID-19 pandemic based on medical information fusion. *Information Fusion.* 2021;73:11-21.
- **Pezoulas VC, Kourou KD, Mylona E, Papaloukas C, Liontos A, Biros D, et al.** ICU admission and mortality classifiers for COVID-19 patients based on subgroups of dynamically associated profiles across multiple timepoints. *Computers in Biology and Medicine.* 2022;141:105176.
- **Han L, Shen P, Yan J, Huang Y, Ba X, Lin W, et al.** Exploring the Clinical Characteristics of COVID-19 Clusters Identified Using Factor Analysis of Mixed Data-Based Cluster Analysis. *Frontiers in Medicine.* 2021;8.
- **Rodríguez A, Ruiz-Botella M, Martín-Loeches I, Jimenez Herrera M, Solé-Violan J, Gómez J, et al.** Deploying unsupervised clustering analysis to derive clinical phenotypes and risk factors associated with mortality risk in 2022 critically ill patients with COVID-19 in Spain. *Critical Care.* 2021;25(1):63.

Integrantes

- **Andrés Gaviria**
- **Luis Daza**
- **Eniver Pino**

- **Denisse Trujillo**