

Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Postgrado
Introducción a la Minería de Datos



Proyecto Parte 2

Glen Abra-ham Calel Robledo
999015052

Introducción

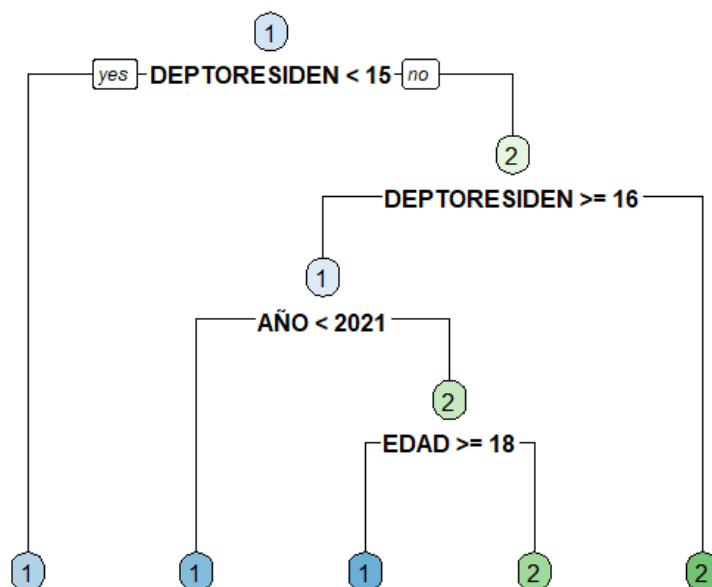
En el curso de introducción a la minería de datos se ha propuesto con el siguiente proyecto realizar la carga y limpieza de varias bases de datos de una misma fuente de datos, los datos son obtenidos desde el Instituto Nacional de Estadística de Guatemala, este organismo se encarga de recolectar, elaborar y publicar estadísticas oficiales con relación a distintos tipos de investigaciones que se realiza en conjunto con otras organizaciones. Para esta ocasión, el tema a evaluar serán las consultas externas que se han realizado en los servicios hospitalarios desde el año 2002 hasta el 2009. Después de haber hecho este proceso se aplicarán algoritmos que encuentren reglas de asociación de datos, de tal manera que se pueda realizar un análisis a las reglas que se encontraron con el fin de identificar patrones llamativos y posteriormente utilizar algoritmos de predicción para determinar si es posible realizar dicha acción utilizando ciertas variables.

Arboles de decisión

Predicción No.1

Para la obtención de esta predicción se analizó el set de datos que abarca los años de 2022 hasta el año 2015, filtrando la causa de enfermedad leucemia linfoblástica aguda [LLA]. Posteriormente se seleccionaron las variables de año, departamento de residencia, edad, pueblo de pertenencia, periodo de edad y tipo de consulta para determinar el sexo de la persona que puede ser más propensa a padecer dicha enfermedad, a partir de esto se obtuvo un resultado gráfico el cual se observa a continuación.

Predicción de sexo



Gracias a este resultado se puede interpretar que en departamentos los siguientes departamentos Guatemala, El Progreso, Sacatepéquez, Chimaltenango, Escuintla, Santa Rosa, Sololá, Totonicapán, Quetzaltenango, Suchitepéquez, Retalhuleu, San Marcos, Huehuetenango, Quiché es más probable que sea un hombre el afecto, en caso contrario si la persona puede pertenecer a los departamentos de Alta Verapaz, Petén, Izabal, Zacapa, Chiquimula, Jalapa y Jutiapa y es mayor o igual de 16 años es probable que sea mujer, y la última observación se puede determinar por el año ya que a partir del 2022 en adelante si es mayor o igual de 18 años puede ser una mujer la afectada.

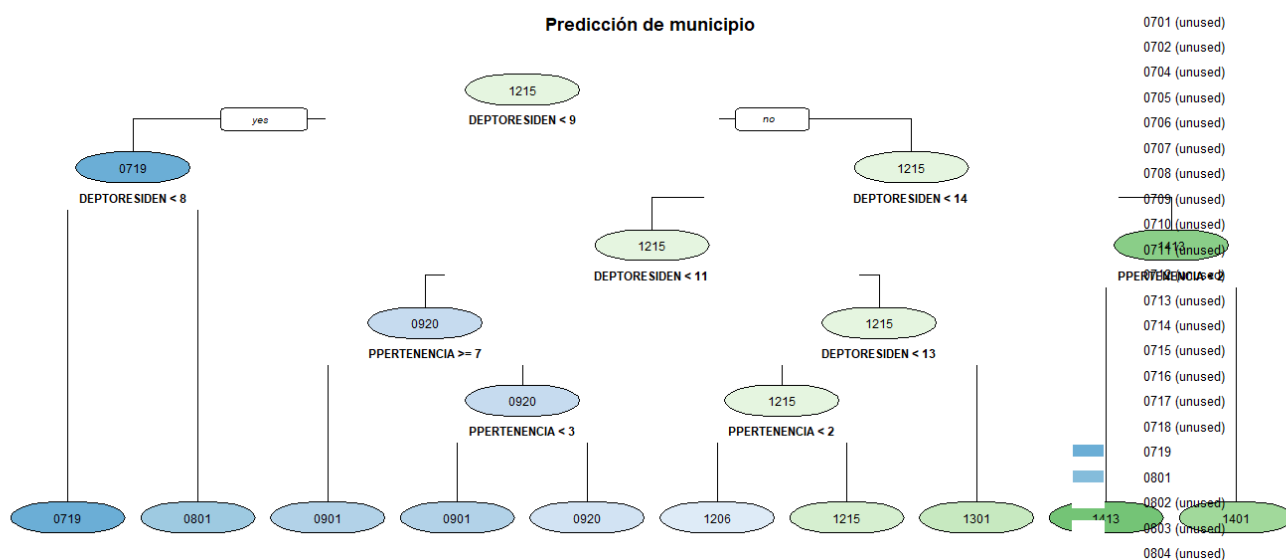
Para probar dicho modelo se realizó la siguiente predicción: determinar el sexo de la persona que puede padecer dicha enfermedad si reside en el departamento de

Guatemala, tiene 18 años, es ladino o mestizo; y como resultado se obtuvo que la persona más propensa con estas características es un hombre.

Según el Manual MSD (2023), indica que esta enfermedad es el cáncer pediátrico más frecuente; también afecta a adultos de todas las edades. Además, también hace énfasis a que la palabra aguda se refiere a que puede progresar rápidamente comenzando con la médula ósea causando síntomas como anemia, trombocitopenia y granulocitopenia. El Manual MSD (2023) indica que el paciente debe ser tratado con 4 fases generales de la quimioterapia.

Predicción No.2

Para la obtención de esta predicción se analizó el set de datos que abarca los años de 2022 hasta el año 2015, delimitando geográficamente la región de occidente que conforman los departamentos de Quetzaltenango, Sololá, Huehuetenango, San Marcos, Quiché y Totonicapán, filtrando la causa de estado de embarazo incidental en mujeres. Posteriormente se seleccionaron las variables edad, pueblo de pertenencia, periodo de edad y departamento de residencia para determinar el municipio donde se podría evitar dicha situación, a partir de esto se obtuvo un resultado gráfico el cual se observa a continuación.



Gracias a este resultado se puede interpretar que, en los departamentos de Guatemala, El Progreso, Sacatepéquez, Chimaltenango, Escuintla, Santa Rosa, Sololá el municipio es Santiago Atitlán, si fuera Totonicapán el municipio es Totonicapán. Otra observación se da con el pueblo de pertenencia, si el pueblo de pertenencia es maya o garífuna y la mujer reside en el departamento de Quetzaltenango o Suchitepéquez, los municipios pueden ser Quetzaltenango o

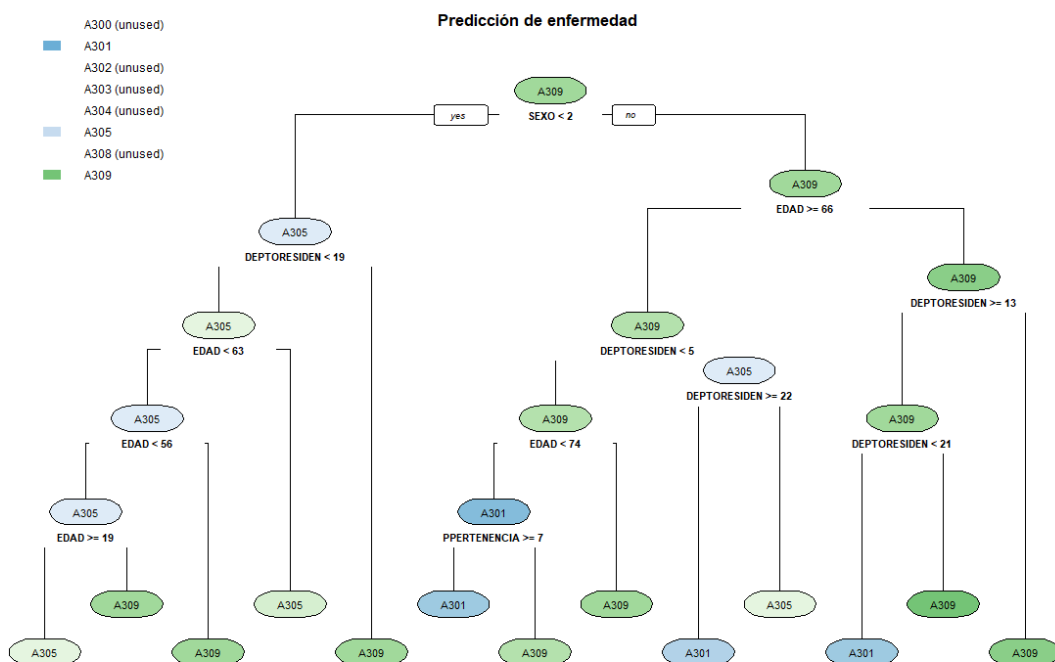
Coatepeque. Y la última observación si reside en San Marcos, el municipio es Concepción Tutuapa si el pueblo de pertenencia es maya, en caso contrario el otro municipio es Malacatán.

Para probar dicho modelo se realizó la siguiente predicción: una mujer de 20 años que sea ladina y que resida en el departamento de Quetzaltenango, el municipio que tiene el problema de embarazos incidentales es Coatepeque.

La BBC (2020) publicó el siguiente artículo *Guatemala, el país centroamericano donde 14 niñas quedan embarazadas toso los días*, según este artículo al menos 5133 niñas de entre 10 y 14 años resultaron embarazadas en el año 2019. Debido a este análisis es muy probable que niñas hayan sido víctimas de este problema que afecta la sociedad teniendo en cuenta que el análisis indica que fueron menores de 24 años.

Predicción No.3

Para la obtención de esta predicción se analizó el set de datos que abarca los años de 2022 hasta el año 2015, teniendo en cuenta que delimitaron las siguientes causas: Lepra [enfermedad de Hansen], Lepra indeterminada, Lepra tuberculoide, Lepra tuberculoide limítrofe, Lepra limítrofe, Lepra lepromatosa limítrofe, Lepra lepromatosa, Otras formas de lepra y Lepra, no especificada. Posteriormente se seleccionaron las variables de edad, pueblo de pertenencia, sexo y departamento de residencia para determinar qué tipo de lepra puede sufrir una persona, a partir de esto se obtuvo un resultado gráfico el cual se observa a continuación.



Por medio de este gráfico se puede interpretar que para los hombres que residen en los departamentos de Guatemala, El Progreso, Sacatepéquez, Chimaltenango, Escuintla, Santa Rosa, Sololá, Totonicapán, Quetzaltenango, Suchitepéquez, Retalhuleu, San Marcos, Huehuetenango, Quiché, Baja Verapaz, Alta Verapaz, Petén e Izabal que sean menores de 19 años pueden sufrir de lepra lepromatosa, si son mayores o igual de 19 años y menores de 56 pueden sufrir de lepra no especificada, o si tienen 55 años y son menores de 63 pueden sufrir de lepra no especificada. Otra observación se da en personas mayores o iguales de 66 años que residen en los departamentos de Huehuetenango, Quiché, Baja Verapaz, Alta Verapaz, Petén, Izabal, Zacapa o Chiquimula pueden sufrir de lepra tuberculoide o que residan en los departamentos de de Guatemala, El Progreso, Sacatepéquez, Chimaltenango, Escuintla, Santa Rosa, Sololá, Totonicapán, Quetzaltenango, Suchitepéquez, Retalhuleu, San Marcos puede sufrir de lepra no especificada.

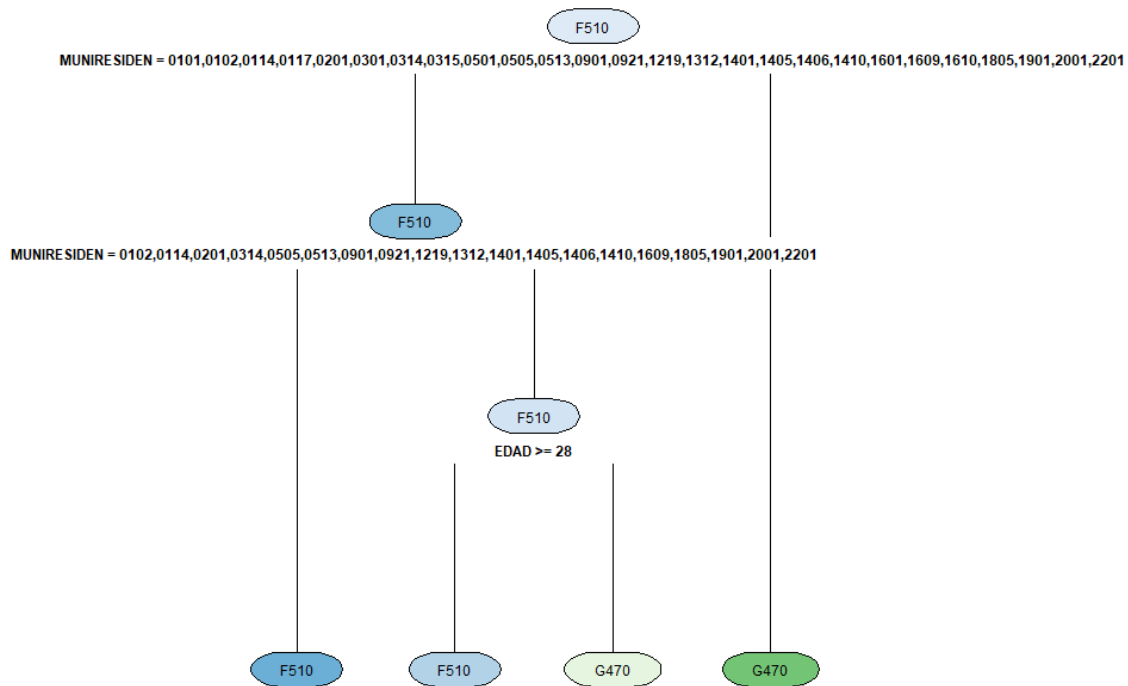
Para probar dicho modelo se realizó la siguiente predicción: una persona de 68 años que sea ladino/mestizo y resida en el departamento de Jalapa se obtuvo como resultado que puede padecer de lepra lepromatosa.

De acuerdo con la información proporcionada por el IGSS (2024), antes de 1940 la lepra se consideraba como una enfermedad incurable que causaba una muerte lenta, pero hoy en día tiene cura. Esta institución recomienda prestar atención a las lesiones que no se curan con tratamientos convencionales, ya que ese es el primer síntoma que puede indicar que la persona puede tener esta enfermedad.

Predicción No.4

Para la obtención de esta predicción se analizó el set de datos que abarca los años de 2022 hasta el año 2015, teniendo en cuenta las casas tales como Trastornos del inicio y del mantenimiento del sueño [insomnios] e Insomnio no orgánico. Posteriormente se seleccionaron las variables de edad, sexo y municipio de residencia para determinar cualquiera de los tipos de insomnios, a partir de esto se obtuvo un resultado gráfico el cual se observa a continuación.

Predicción de trastorno de sueño



Gracias a este gráfico se puede determinar que las personas que residen en el municipio de Guatemala, Santa Catarina Pinula, Amatitlán, Petapa, Guastatoya, Antigua Guatemala, Alotenango, San Antonio Aguas Calientes, Escuintla, Masagua, Nueva Concepción, Quetzaltenango, Génova, San Pablo, La Democracia, Santa Cruz del Quiché, Chajul, Chichicastenango, Cunén, Cobán, San Pedro Carchá, San Juan Chamelco, Los Amates, Zacapa, Chiquimula y Jutiapa pueden sufrir de insomnio, y si se tiene en cuenta la edad que sean menores de 28 años pueden sufrir insomnio no orgánico.

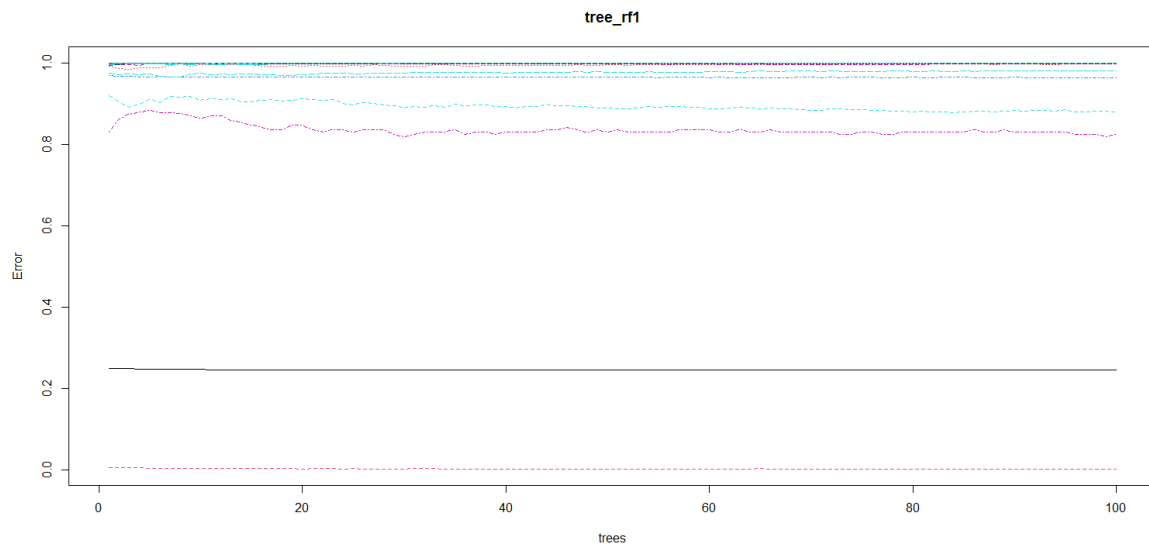
Para probar dicho modelo se realizó la siguiente predicción: una persona de 40 años que reside en el municipio de Nueva Concepción sin importar su sexo es propensa a tener insomnio no orgánico.

El insomnio es un trastorno que afecta la calidad de vida de las personas, según el artículo *Insomnio en adultos y ancianos: enfoque en atención primaria* (Lizeth X. Morales-Ceballos, 2024) indica que el insomnio es un trastorno prevalente el cual debe ser tratado con el control de comorbilidades, higiene del sueño y terapia cognitivo-conductual, en caso de dejar el problema la persona puede padecer de síntomas como malestar físico, fatiga, baja motivación, estado de ánimo irritable entre otros.

Bosques Aleatorios

Predicción No.1

Para la obtención de esta predicción se analizó el set de datos que abarca los años de 2022 hasta el año 2015, teniendo en cuenta las enfermedades como Miopía y Astigmatismo. Posteriormente se seleccionaron las variables de edad, pueblo de pertenencia, tipo de enfermedad y el sexo, a partir de esto se obtuvo el siguiente gráfico con todos los árboles el modelo generó.



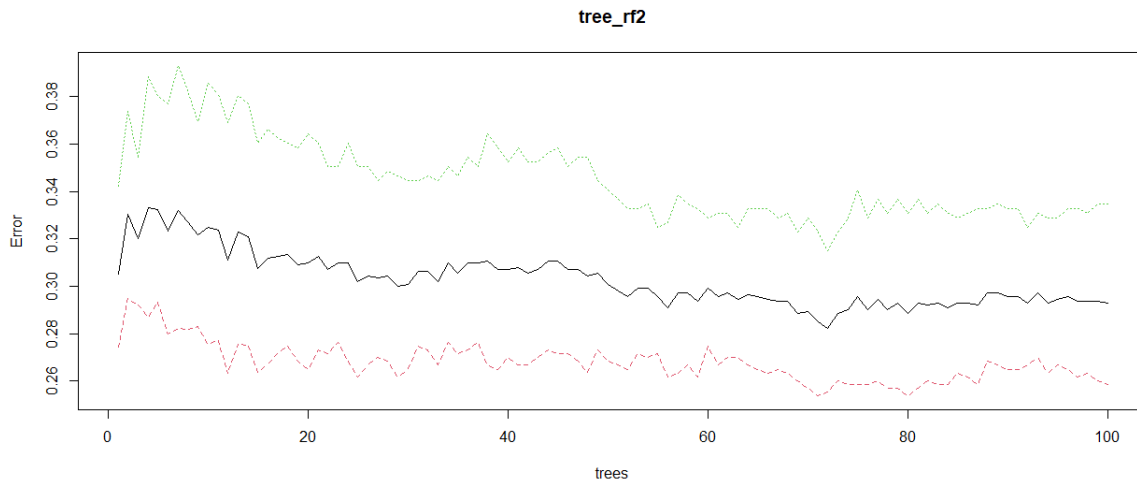
Para probar dicho modelo se realizó la siguiente predicción: un hombre de 18 años que su pueblo de pertenencia es maya y tiene miopía, el modelo arroja como resultado que es muy probable que esta persona resida en Alta Verapaz.

De acuerdo con el artículo *¿Han aumentado los casos de miopía en niños por el uso de pantallas digitales?* (Ola, 2023) durante la pandemia se incrementaron los problemas de miopía en la población infantil en Guatemala debido al confinamiento y la principal causa de esta enfermedad se debe al uso de pantallas digitales debido a que la mayoría de los alumnos tenía que llevar sus cursos en línea. Además, es importante mencionar que no existe una edad como tal una enfermedad visual, pero es recomendable acudir a un oftalmólogo para prevenir este tipo de problemas.

Predicción No.2

Para la obtención de esta predicción se analizó el set de datos que abarca los años de 2022 hasta el año 2015, teniendo en cuenta la enfermedad de Parkinson. Posteriormente seleccionaron las variables de departamento de residencia, pueblo

de pertenencia y edad para determinar el sexo de la persona afectada, a partir de esto se obtuvo el siguiente gráfico con todos los árboles el modelo generó.



Para probar dicho modelo se realizó la siguiente predicción: una persona que reside en el departamento de Escuintla, de pueblo maya y edad de 70 años, el modelo arroja como resultado que una persona con tales características puede sufrir de Parkinson siendo hombre.

La enfermedad de Parkinson se debe a la pérdida gradual de neuronas en el cerebro, teniendo en cuenta que se debe a factores como la edad y la genética, pero recientemente se ha determinado que el ambiente también es un factor crucial que, según el artículo *Pesticidas y párkinson: una conexión fuera de toda duda* (Villapol, 2024) indica que se han hecho estudios para determinar que los insecticidas, fungicidas y herbicidas influyen significativamente en el deterioro de las neuronas. Teniendo en cuenta que Guatemala es un país que tiene grandes proporciones de tierra para la agricultura, es muy importante tener el cuidado de la población que utiliza este tipo de productos.

Bibliografía

- Disorders, N. I. (2023 de Marzo de 2023). *National Institute ON Deafness and Other Communication Disorders*. Obtenido de National Institute ON Deafness and Other Communication Disorders: <https://www.nidcd.nih.gov/es/espanol/trastorno-del-desarrollo-del-lenguaje>
- Emadi, A. (Octubre de 2023). *Manual MSD*. Obtenido de Manual MSD: <https://www.msmanuals.com/es/professional/hematolog%C3%ADa-y-oncolog%C3%ADa/leucemias/leucemia-linfobl%C3%A1stica-aguda-Ila>
- IGSS. (28 de Enero de 2024). *IGSS*. Obtenido de IGSS: <https://www.igssgt.org/noticias/2024/01/28/lepra-una-enfermedad-curable-que-continua-creando-estigmas-que-dificultan-su-diagnostico-y-tratamiento/>
- James Fernandez. (Octubre de 2022). *Manual MSD*. Obtenido de Manual MSD: <https://www.msmanuals.com/es/professional/inmunolog%C3%ADa-y-trastornos-al%C3%A9rgicos/enfermedades-al%C3%A9rgicas-autoinmunitarias-y-otros-trastornos-por-hipersensibilidad/rinitis-al%C3%A9rgica>
- Lizeth X. Morales-Ceballos, G. L.-O.-B.-M. (2024). Insomnio en adultos y ancianos: enfoque en atención primaria.
- Médicos, E. d. (25 de Julio de 2022). *ADA*. Obtenido de ADA: Se denomina hipertensión esencial o hipertensión primaria a la presión arterial alta (más de 140/90 mmHg) sin una causa reconocible.
- National Heart, L. a. (12 de Enero de 2023). *National Heart, Lung, and Blood Institute*. Obtenido de National Heart, Lung, and Blood Institute: <https://www.nhlbi.nih.gov/es/salud/bronquitis>
- Navarra, C. U. (s.f.). *Clinica Universidad de Navarra*. Obtenido de Clinica Universidad de Navarra: <https://www.cun.es/enfermedades-tratamientos/enfermedades/hipertrofia-benigna-prostata>
- Ola, A. L. (2 de Marzo de 2023). *Prensa Libre*. Obtenido de Prensa Libre: <https://www.prensalibre.com/guatemala/comunitario/han-aumentado-los-casos-de-miopia-en-ninos-por-el-uso-de-pantallas-digitales/>
- Villapol, S. (13 de Enero de 2024). *Prensa Libre*. Obtenido de Prensa Libre: <https://www.prensalibre.com/vida/salud-y-familia/pesticidas-y-parkinson-una-conexion-fuera-de-toda-duda/>
- Wallace, A. (14 de Octubre de 2020). *BBC*. Obtenido de BBC: <https://www.bbc.com/mundo/noticias-america-latina-54484133>

Enlace al repositorio

https://github.com/glencalel/data_mining_p2