UNIVERSIDAD AUTONÓMA DE NUEVO LEÓN

Facultad de Ciencias Físico-Matemáticas

Minería de Datos

AVANCE I PROYECTO INTEGRADOR

M.C. Mayra Berrones Reyes

Equipo: 1

Grupo 3, Viernes 18:00 – 21:00 horas

1801925	Alanis Aguirre Arleth
1887833	Delgado Cantú Armando Javier
1931548	Garza Espinosa Omar Alejandro

Séptimo Semestre

30 septiembre 2020

Título de la base de datos **Enfermedades cardiovasculares**

Conjunto de datos de enfermedades cardiovasculares

https://www.kaggle.com/sulianova/cardiovascular-disease-dataset?select=cardio_train.csv

Descripción de los datos

El tipo de datos con los que vamos a trabajar es una tabla conformada por trece columnas, el conjunto de datos se divide de la siguiente manera:

1. Número de identificación del paciente (entero)
2. Edad del paciente en días (entero)
3. Género del paciente (entero)
   donde: 1= mujer
   2= hombre
4. Altura del paciente en cm (entero)
5. Peso del paciente en kg (entero)
6. Presión arterial sistólica (entero)
7. Presión arterial diastólica (entero)
8. Colesterol (entero)
   donde: 1= normal
   2= por encima de lo normal
   3= muy por encima de lo normal
9. Glucosa (entero)
   donde: 1=normal
   2=por encima de lo normal
   3= muy por encima de lo normal
10. Fumar (binario)
11. Ingesta de alcohol del paciente (binario)
12. Actividad física del paciente (binario)
13. Presencia o ausencia de enfermedad cardiovascular (binario)

*En los datos binarios el número 1 es que si realiza la actividad mencionada y número cero no realiza la actividad mencionada.

Justificación del uso de datos

Nos llamó la atención este conjunto de datos porque observamos información muy relevante conforme a la salud de 70,000 personas y que pueden presentar posiblemente una enfermedad cardiovascular. Buscamos prevenir, alertar y dar recomendaciones sobre los factores que aumentan el riesgo de tener alguna enfermedad cardiovascular.

El uso de este conjunto de datos nos beneficia a entender cómo se comportan los ítems que influyen a que una persona presente enfermedad cardiovascular o no, prevenir a la sociedad sobre los factores que influyen para desarrollar estar enfermedad, y así crear un plan estratégico para el

sector salud que brinde información sobre las actividades que impactan al paciente de manera positiva o negativa a la salud.

Planteamiento del problema

Se ha observado que enfermedades como la diabetes, hipertensión y obesidad son factores de riesgo para enfermedades cardiovasculares. Revisando bibliografía y estadística, individuos que presentan una o varias de estas enfermedades tienen mayor incidencia en mortalidad. Al igual de que estas enfermedades complican el estado de salud de las personas. Hoy más que nunca es necesario detectar la población en riesgo ya que esto ayudaría a que las estrategias de prevención de enfermedades del sector salud sean eficientes y a que el gobierno pueda administrar el presupuesto de mejor manera.

Objetivo Final

Crear una herramienta tecnológica precisa para detectar el riesgo de enfermedad cardiovascular en la población apoyado en información clínica y de laboratorio de la población a tratar (base de datos de enfermedades cardiovasculares).

Nuestro objetivo principal es brindar información al sector salud sobre la población y la probabilidad de la mayoría de los factores (peso, edad, presión arterial sistólica y diastólica, cigarro, alcohol, colesterol, glucosa y actividad física) que influyen a desarrollar enfermedades cardiovasculares así ellos podrán utilizar estos resultados para elaborar programas de salud que beneficien a la población, así como al gobierno a administrar mejor sus recursos.

Como objetivos secundarios tenemos el encontrar alguna relación entre las variables de manera particular que influyan en la presencia de la enfermedad, y así ver puntualmente que factores afectan significativamente a presentar dicha enfermedad.

Planeación de la herramienta

Para la herramienta que realizaremos contemplamos utilizar tres técnicas de Minería de Datos, dado que nuestro objetivo es encontrar los factores más influyentes, decidimos tomar Clustering y Reglas de Asociación. La tercera técnica es Regresión, ésta se asocia más con nuestro objetivo secundario ya que determina la influencia de las variables.

Usaremos Clustering porque queremos identificar y agrupar los factores que presentan características similares para ver cuáles influyen a desarrollar enfermedad cardiovascular. Reglas de asociación la utilizaremos para encontrar la relación entre los factores, y descubrir los hechos comunes dentro de este conjunto de datos. Ambas son de ayuda para identificar patrones en las variables que influyen en desarrollar o no la enfermedad. Mientras que la técnica de Regresión nos ayudaría a encontrar alguna relación entre las variables con la presencia o no de la enfermedad y determinar cuál es la que influye con mayor intensidad.