**Identificación del problema:**

**Descripción del contexto problemático (causas y síntomas)**

Las enfermedades más comunes relacionadas a problemas en el corazón son la Angina, Infarto de miocardio, la Hipertensión entre otras (Henríquez, 2018), Estas enfermedades coronarias deben ser tratadas a tiempo de lo contrario esto puede llevar graves consecuencias como la creación de más enfermedades relacionadas y en el peor caso la muerte. En la ciudad de Cleveland en los Estados Unidos se identificó la importancia de esto y se realizó un estudio a una muestra de 304 personas para identificar si estas contaban con problemas en el corazón.

A pesar de que los datos se encuentran registrados de una manera consistente y cuentan con información valiosa estos datos no se encuentran estructurados por lo cual si se quisiera realizar un *análisis a la antigua* sería necesario tener un amplio equipo encargado de leer y analizar cada uno de estos casos registrados de manera manual en los que se incluyen más de 13 parámetros, en otras palabras ¡sería necesario analizar 4.256 datos uno por uno! Por si esto fuera poco, suponiendo que esta labor tiene éxito, sería necesario también realizar un análisis profundo de los datos ya que no basta con solo saber el contenido de la muestra si no que la verdadera importancia radica en poder inferir si una persona tiene problemas cardiacos, sería muy *engorroso* para un médico buscar en el *dataset* un paciente que tenga las mismas características y mirar si este tendrá problemas en el corazón o no, es más, es muy probable simplemente no lo encuentre y que este se equivoque en la búsqueda. Estas son labores que de cierto modo son inalcanzables si se hacen manualmente pero un *software* puede resolver en unos pocos segundos, la idea es que mediante la carga de datos y el Machine Learning se pueda facilitar la labor de los médicos a la hora de determinar si una persona tiene o puede tener problemas en el corazón, una labor que debe realizarse con sumo cuidado debido al sector en el que se encuentra el proyecto.

**Identificación de necesidades:**

**Definición del problema:**

La ciudad de Cleveland ha realizado una investigación sobre personas que tienen y no tienen problemas en el corazón, a cada persona se le ha recopilado la edad, sexo, dolor de pecho, presión arterial en reposo, colesterol en la sangre, resultado de examen de azúcar en sangre en ayunas, resultado del electro cardiograma en reposo, frecuencia cardiaca máxima en reposo, si tiene angina inducida por ejercicio, ejercicio relativo al descanso, *slope*, numero de vasos sanguíneos principales coloreados por fluoroscopia, Talasemia y el registro de si tiene problemas en el corazón o no. Se requieren ordenar, visualizar e inferir si una persona tendrá problemas en el corazón de manera concisa

**Recopilación de la información necesaria**

El dolor de pecho es un síntoma común que se presenta en personas con un mal flujo sanguíneo al corazón y que puede significar un riesgo de ataque al corazón (Chen, s.f.) si se realiza un examen este pude dar como resultado 3 valores.

1. Angina
2. Asintomático
3. Anormal

El electrocardiograma es un examen que registra la actividad eléctrica del corazón y permite detectar si una persona ha tenido algún infarto o si tiene arritmias cardiacas este puede generar alguno de los siguientes resultados (Mañero, 2019)

1. Normal
2. Anormal
3. Hipertenso

La angina inducida por ejercicio es si una persona presenta dolor de pecho al someterse a alguna actividad física (Chen, s.f.)

*Slope*: Es una pendiente que representa una tolerancia al ejercicio físico (R S Finkelhor, 1986)

Un árbol de decisión aplicado en el machine learning es conjunto de datos que se forman (Valga la redundancia) en forma de árbol dependiendo a una función que el programador asigna, este conjunto de datos tiene muchas utilidades, sus dos principales son poder determinar si un set de datos de entrada pertenecen a una determinada clase esta puede ser binaria o \_\_\_\_ , en ciertos casos los que el árbol no sea muy grande puede permitir visualizar de manera intuitiva cómo está funcionando la clasificación de datos en un *dataset*.

La regresión en el machine learning es una técnica en la cual el resultado es un valor numérico en un conjunto infinito de soluciones como estimar que cantidad de productos se va a vender en un supermercado (Heras, 2020)