**Objetivos**

**Terminales del Curso**

OT1. Desarrollar programas de computador para solucionar problemas de ingeniería de mediana complejidad que involucren el análisis de volúmenes significativos de datos.

OT2. Evaluar los resultados del análisis de datos utilizando teoría de probabilidades e inferencia estadística.

OT3. Evaluar programas de computador utilizando técnicas de análisis de algoritmos y estructuras de datos.

OT4. Sustentar apropiadamente los resultados de cada una de las etapas del proceso de desarrollo de soluciones de ingeniería de forma escrita y oral.

**Enunciado**

Su grupo de trabajo, cuyo tamaño máximo son 4 personas, debe elegir un conjunto de datos (dataset) con base en el cual deben definir un problema a solucionar. El problema debe ser abordado utilizando el método de la ingeniería.

El proyecto propuesto debe ser un problema de **análisis de información**, concretamente debe tratar sobre **clasificación**. En la solución deben utilizar la técnica de **árboles de decisión**, en la que su equipo de trabajo debe desarrollar (en C# sobre .NET) la estructura de datos y los algoritmos para la construcción de dicho árbol, es decir, en esta parte de la solución no se puede utilizar librerías externas de análitica de datos.

**Fuentes de Datasets para Clasificación**

A continuación se presentan algunas fuentes de datasets interesantes donde **pueden** (pero no tienen que) buscar el dataset que utilizarán en su proyecto:

**Fuente 1.** <http://archive.ics.uci.edu/ml/index.php>

**Fuente 2.** <https://www.kaggle.com/datasets>

Fuente 3. <https://datos.gov.co/>

Fuente 4. <https://www.dane.gov.co/>

**Textos de Referencia**

**Libro 1.** [The Data Science Design Manual](https://bit.ly/2FO78P0)

**Libro 2:** [An Introduction to Machine Learning](https://bit.ly/2FYwdGu)

**Entregables**

**Entrega 1.** Fecha de Entrega Máxima: Viernes 16 de Abril de 2021, 23:55.

1. Requerimientos Funcionales y No Funcionales de todo el proyecto.
2. Primeras 3 etapas del Método de la Ingeniería.
3. Iteración 1. Desarrollo del módulo de exploración de datos. Reportes parametrizados tabulares (1 reporte por cada columna) y gráficos (5) sobre los datos.
   1. Diagrama de Clases de esta iteración.
   2. Diagrama de Objetos con base en el diagrama de clases anterior.
   3. Implementación en C# sobre .NET.
4. En el repositorio de proyecto en GitHub. Todos los integrantes deben ser colaboradores del repositorio y cada uno debe tener al menos 10 commits en total (espaciados a lo largo de mínimo 1 semana).
5. El readme.md del repositorio debe explicar brevemente (en inglés) de qué se trata el proyecto. Deben enlazar los archivos que documentan el proyecto (en formato pdf) y deben especificar las condiciones técnicas del mismo (lenguaje, sistema operativo y ambiente de desarrollo).
6. Video explicativo de la entrega. En el video deben presentar todos los integrantes del equipo. En el momento en que una persona se encuentre hablando debe visualizarse su video en una miniatura. En el video deben presentar de forma breve la documentación solicitada, haciendo énfasis en los elementos más importantes de esta entrega, al igual que una demostración de la implementación. Duración máxima del video: 5 minutos.

**Descripción del problema**

Una reconocida clinica de cardiologia en cali, necesita saber si sus pacientes son propensos a sufrir un ataque cardíaco, pero dado a que este proceso suele tomar mucho tiempo, el cual puede ser vital para el paciente, decidieron que lo mejor era contratar un grupo de expertos en programación para saber si era posible automatizar este proceso. Para ello decidieron reunir un historial de datos de los pacientes a las que se les ha hecho pruebas para saber si eran o no propensos a sufrir ataques cardiacos, los datos del paciente que incluía este historial eran la edad, el género, el tipo de dolor de pecho (angina típica, angina atípica, dolor no anginoso, asintomático), presión arterial, el nivel de colesterol, el nivel de azúcar en la sangre, si ha sufrido una angina (dolor en el pecho), los resultados de un electrocardiograma en reposo, la frecuencia cardiaca máxima y el resultado de si es propenso o no a sufrir ataques cardiacos. La clínica ordenadamente ha recopilado esta información en un dataset y ha contratado al mejor equipo de ingenieros de software de la ciudad de cali para este trabajo.

Aparte de mostrar el resultado la clinica tambien solito que el programa permitiera visualizar todos los datos en una tabla los cuales pueden ser filtrados por el tipo de angina, también se necesita que se grafiquen algunos datos para que sea más fácil de visualizar y también comprender los factores más importantes del dataset, estos gráficos deben ser: la cantidad de mujeres y hombres que son propensos a sufrir ataques, los pacientes que han tenido angina inducida por ejercicio contra los que no, rangos de edad de los pacientes, los diferentes tipos de dolor en el pecho y por último un gráfico sobre el nivel de colesterol de los pacientes que son propensos contra los que no a sufrir ataques al corazón.

Dado a que no son muy buenos con la tecnología, les piden que los datos de los pacientes los puedan ingresar de dos formas, la primera es mediante una base de datos con el mismo formato que la base presentada anteriormente pero sin la columna de la solución, y la otra opción es en la que puedan ir agregando cada registro manualmente de manera intuitiva y cuando esté listo poner a correr el programa y este les muestre por pantalla el resultado pero que también les permite almacenarlo en el ordenador.

**Requerimientos funcionales**

El sistema debe estar en la capacidad de :

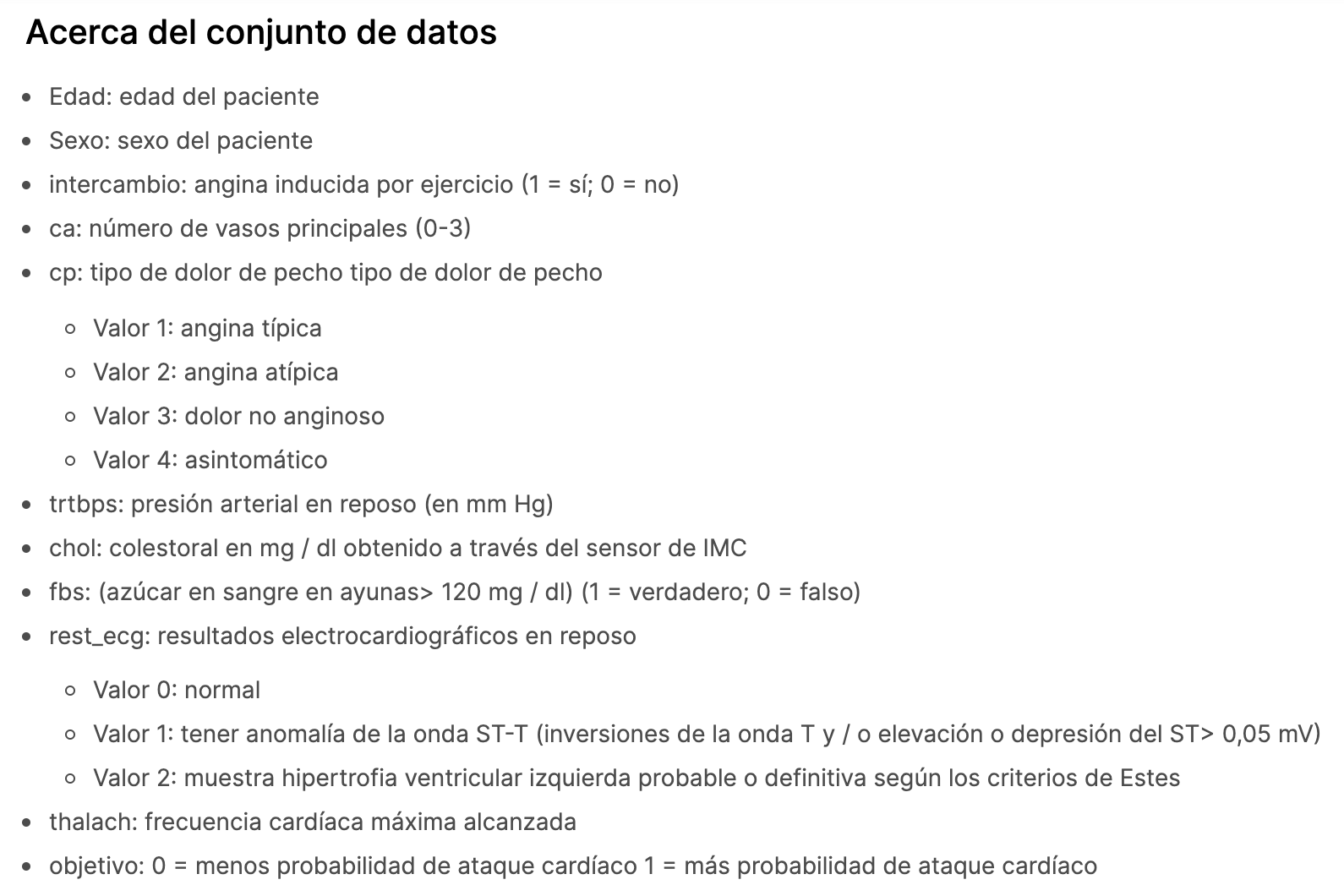
* Predecir con los datos de un paciente, si es probable que este sufra un ataque cardiaco o no.
* Filtrar un reporte por paciente donde me muestre los datos organizados por un tipo de dolor en el pecho, este debe ser mediante un combobox de opciones que me muestre los 4 valores:
  + Valor 1: angina típica
  + Valor 2 : angina atipica
  + Valor 3: dolor no anginoso
  + Valor 4: asintomático
* Visualizar un gráfico que muestre:
  + la cantidad de mujeres y hombres que son propensos a sufrir un ataque al corazón.
  + Los pacientes que han tenido una angina inducida por ejercicio contra los que no han sufrido dicho padecimiento.
  + Los diferentes rangos de edad de los pacientes
  + Un gráfico de pastel que muestre la cantidad de los pacientes que han tenido un dolor en el pecho, clasificando esto por su tipo.
  + Un gráfico de líneas que indiquen los niveles de colesterol de cada paciente.
* Leer el dataset de los para entrada de datos en un formato específico.
* Escribir un archivo tipo csv para mostrar la solución del problema.
* Una interfaz intuitiva para el manejo del programa, esta debe permitir desarrollar al máximo las funcionalidades del sistema.
* El sistema debe estar en la capacidad de realizar los siguientes registros:
  + Agregar registros de pacientes
  + Borrar registros
  + Actualizar registros
* El sistema debe permitir un reporte de los datos en los siguientes tipos:
  + Categorico:seleccionar una opcion del combo box para identificar, el sexo, los tipos de dolores en el pecho, la angina inducida por ejercicio,, resultados del electrocardograma, el auzcar en la sangre en ayunas.
  + Numérico: un rango de valores para filtrar los datos del dataset: por edad, nivel de colesterol, presión arterial en reposo, la frecuencia máxima cardiaca, números de vasos principales
  + Cadena:
* Clasificar los datos

Requerimientos no funcionales:

Diagrama de clases

Clases:

Clinica:

Paciente : 

Método de la ingeniería

Identificación: Se requiere predecir teniendo en cuenta los datos de un paciente si este es propenso a sufrir un ataque al corazón.

Necesidades:

* Identificar la historia clínica del paciente
* Identificar los datos irrelevantes del dataset
* Predecir una variable objetivo
* Identificar la variable objetivo

Recopilacion de informacion:

Ataque al corazón:

Árboles de decisión

Búsqueda de soluciones creativas