



Departamento de TIC
Algoritmos y Estructuras de
Datos

PROYECTO INTEGRADOR

SISTEMA DE CARACTERIZACIÓN DE RIESGOS DE SUICIDIO A NIVEL GLOBAL

MÉTODO DE INGENIERÍA

POR:

ANDREA NÚÑEZ

CAMILO ESCOBAR

DANNA GARCÍA

CAMILO GUTIÉRREZ

DIRIGIDO A:

JUAN MANUEL REYES.

INGENIERO DE SISTEMAS Y COMPUTACIÓN.

UNIVERSIDAD ICESI

SANTIAGO DE CALI

COLOMBIA

2021

FASE 1: IDENTIFICACIÓN DEL PROBLEMA

Definición del problema

Actualmente, mientras se realiza cualquier actividad cotidiana, sólo en Colombia, 7 personas se habrán quitado la vida y más de 200 personas lo habrán intentado; según datos oficiales de la Organización Mundial de la Salud (OMS). El suicidio es la primera causa de muerte entre las personas, sobre todo en adolescentes, a nivel mundial y también es una de las principales causas de muerte no natural en nuestro país. Históricamente, se han registrado los datos de las personas que se han suicidado, haciendo una distinción de la generación según la época en la que se encontraban, por ejemplo, se puede encontrar la cantidad de *millennials* que se han suicidado y el lugar donde ocurrió, permitiendo sacar conclusiones sobre las causas de estos suicidios y si están relacionados a los sucesos más relevantes del momento. De esta manera, se requiere un software que permita conocer qué tan alta fue la tasa de suicidios de algún país dada una época y una generación, con el objetivo de asociar estos eventos a acontecimientos históricos que ocurrieran en ese instante.

Descripción: Data set que registra la tasa de suicidios a nivel mundial desde 1985 a 2016. El data set consta de 12 columnas y 27820 registros.

Alojamiento: <https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016>

Referencias.

United Nations Development Program. (2018). Human development index (HDI). Retrieved from <http://hdr.undp.org/en/indicators/137506>

World Bank. (2018). World development indicators: GDP (current US\$) by country:1985 to 2016. Retrieved from <http://databank.worldbank.org/data/source/world-development-indicators#>

[Szamil]. (2017). Suicide in the Twenty-First Century [dataset]. Retrieved from <https://www.kaggle.com/szamil/suicide-in-the-twenty-first-century/notebook>

World Health Organization. (2018). Suicide prevention. Retrieved from http://www.who.int/mental_health/suicide-prevention/en/

Variable objetivo: Nivel de riesgo de suicidio en población millennial de un país en un año específico (Bajo riesgo, riesgo moderado, alto riesgo).

Identificación de necesidades y síntomas

- Encontrar qué tan alto es el ratio de suicidios de un país según una época dada.

- Lograr identificar el nivel de riesgo de suicidio de cada una de las generaciones.
- Elaborar informe sobre los factores estudiados y técnicas con implementación propia e implementación externa

Comentado [CC1]: Revisar clasificación de requerimientos (Enfoque en informes hechos por nosotros)

Requerimientos funcionales.

1. Cargar el data set Suicide Rates Overview 1985 to 2016 dado en un archivo .csv al sistema.
2. Generar un reportes parametrizados por cada columna del data set Suicide Rates Overview 1985 to 2016.
3. Producir 5 gráficos estadísticos sobre todo el reporte de Suicide Rates Overview 1985 to 2016.
4. Caracterizar la tasa de riesgo de un país según un año dado con un enfoque generacional variable.
5. Implementar el módulo de ambiente de ejecución de experimentos partiendo de las variables dadas en el data set Suicide Rates Overview 1985 to 2016.
6. Generar reporte con análisis estadístico sobre las variables de salida del experimento (ANOVA, Poison, etc.).

Requerimientos no funcionales.

1. La implementación debe realizarse sobre C# haciendo uso del Framework .NET.
2. La primera caracterización de riesgos debe hallarse mediante un árbol de decisión de implementación propia.
3. La segunda caracterización de riesgos debe hallarse mediante un árbol de decisión utilizando una librería externa de C#.

FASE 2: RECOPIACIÓN DE INFORMACIÓN NECESARIA

Con el objetivo de solucionar este problema y obtener un correcto funcionamiento del software, es esencial conocer y entender algunos conceptos y herramientas fundamentales que estén relacionadas con el cumplimiento de los requerimientos funcionales previamente mencionados.

Conceptos del dominio de conocimiento del problema:

- **Suicidio:** Es el acto de quitarse deliberadamente la propia vida. El comportamiento suicida es cualquier acción que pudiera llevar a una persona a morir, como tomar una sobredosis de medicamentos o estrellar un automóvil a propósito. ([Fuente](#))
- **Ratio:** Una ratio es el cociente entre dos magnitudes que están relacionadas. El objetivo es poder establecer cálculos y realizar comparaciones a través de este instrumento. Los ejemplos son muy utilizados en economía, sobre todo en el ámbito financiero. ([Fuente](#))
- **Riesgo:** El riesgo es la probabilidad de que una amenaza se convierta en un desastre. La vulnerabilidad o las amenazas, por separado, no representan un peligro. Pero si se

juntan, se convierten en un riesgo, o sea, en la probabilidad de que ocurra un desastre. ([Fuente](#))

- **Población:** La población es el conjunto de personas o animales de la misma especie que se encuentran en un momento y lugar determinado. ([Fuente](#))
- **Generación:** Conjunto de personas, generalmente dedicadas al arte o a la ciencia, cuya obra tiene características comunes y han vivido los mismos acontecimientos históricos. ([Fuente](#))
 - ⊗ Baby Boomers: nacidos entre 1946 y 1964.
 - ⊗ Generación X: nacidos entre 1965 y 1979.
 - ⊗ Millennials o Generación Y: nacidos entre 1980 y 1999.
 - ⊗ Generación Z: nacidos a partir del 2000.

Conceptos del dominio de conocimiento del software:

- **Data set:** es una colección de datos habitualmente tabulada que contiene los valores para cada una de las variables. ([Fuente](#))
- **Librería:** es uno o varios archivos escritos en un lenguaje de programación determinado, que proporcionan diversas funcionalidades. ([Fuente](#))
- **C#:** Es un lenguaje de programación multiparadigma desarrollado y estandarizado por la empresa Microsoft como parte de su plataforma .NET, que después fue aprobado como un estándar por la ECMA e ISO. C# es uno de los lenguajes de programación diseñados para la infraestructura de lenguaje común. ([Fuente](#))
- **Árbol de decisión:** Un árbol de decisión es un mapa de los posibles resultados de una serie de decisiones relacionadas. Permite que un individuo o una organización comparen posibles acciones entre sí según sus costos, probabilidades y beneficios. Se pueden usar para dirigir un intercambio de ideas informal o trazar un algoritmo que anticipe matemáticamente la mejor opción. ([Fuente](#))
- **Framework:** Es un conjunto estandarizado de conceptos, prácticas y criterios para enfocar un tipo de problemática particular que sirve como referencia, para enfrentar y resolver nuevos problemas de índole similar. ([Fuente](#))
- **.NET:** Framework desarrollado por Microsoft y estandarizado en el desarrollo de software sobre C#. ([Fuente](#))

FASE 3: BUSQUEDA DE SOLUCIONES CREATIVAS

Mediante la estrategia de BrainWriting se obtienen las siguientes soluciones potenciales enfocados a cada ítem

Las calificaciones están separadas por comas y van del 1 al 3, donde 1 es mala idea, 2 es idea regular y 3 es buena idea.

- Carga el data set:
 - Idea 1: Cargar el archivo mediante archivos de texto.

Comentado [CC2]: Debatir si se agrega información de herramientas (GIT, VISUAL, ETC)

- Idea 2: Mediante una herramienta propia de .NET (StreamReader) cargar cada línea del archivo .csv.
 - Idea 3: Cargar la base de datos mediante un algoritmo programado en C# que permita leer línea por línea el documento y almacenarlos.
- Almacenar registros:
 - Idea 1: Almacenar los datos en hash table.
 - Idea 2: Utilizar SQL Server para almacenar los datos.
 - Idea 3: Emplear una estructura de alguna librería que esté incluida en el lenguaje de programación C#.
- Generar reporte por columnas:
 - Idea 1: generar el reporte en un table view.
- Generar gráficos estadísticos:
 - Idea 1: Utilizar las librerías de .NET para generar gráficos de barras, pastel y dispersión.
 - Idea 2: Valerse de una herramienta de generación de gráficos brindada por un software tal como Excel.
- Clasificar la caracterización de riesgo de suicidio:
 - Idea 1: Realizar la clasificación de los riesgos por países haciendo uso de un árbol de decisión.
 - Idea 2: Realizar la clasificación de los riesgos por países haciendo un árbol de regresión
- Ambiente experimental:
 - Idea 1: Modelar en el sistema un módulo con herramientas básicas de análisis experimental como “Configuración de ambiente experimental, ejecución del ambiente y pruebas y reporte de resultados de las pruebas realizadas”.
 - Idea 2: Incluir dentro del sistema a desarrollar una serie de métodos con la capacidad de interpretar las entradas de los tratamientos con los cuales se va a realizar el experimento y que estos mismos generen las salidas correspondientes al resultado de la variable estudiada y con esta información sacar las conclusiones pertinentes.
- Análisis estadísticos:
 - Idea 1: Una vez generados y exportados los reportes de las pruebas a un archivo .csv valerse de herramientas ofrecidas por R Studio para realizar los análisis estadísticos avanzados que se solicitan.
 - Idea 2: Ya generados los datos y los respectivos reportes, analizar los datos a través del programa estadístico informático SPSS, con el cual se realizan estudios avanzados en este campo.
- Informes:
 - Idea 1: Una vez obtenidos los análisis estadísticos se procede a revisar y hallar conclusiones relevantes de forma manual en Excel.
 - Idea 2: Realizar una recopilación de los análisis y reportes obtenidos y con estos generar un documento en Word donde se especifiquen las conclusiones de los resultados.
 - Idea 3: Realizar un análisis estadístico con los algoritmos mencionados anteriormente e implementar estos resultados en una en una table view donde se puede personalizar las variables que deseamos observar.

FASE 4: PASAR DE LA IDEA PRINCIPAL AL DISEÑO PRELIMINAR

En este paso hacemos un filtro de soluciones más potenciales enfocados a cada ítem, escogiendo la mejor opción:

- Carga del dataset:
 - Decisión de la solución A: Se considera que la mejor alternativa para cargar los datos al sistema es el uso de las herramientas propias de .NET para leer el archivo .csv. Las otras dos ideas se descartan ya que suponen más tiempo de implementación y menos eficiencia de recursos.
- Almacenar registros:

Aproximaciones:

 - Idea 3: Se considera que la mejor alternativa para almacenar los registros en el sistema es mediante el empleo de una librería existente de .NET, pues esto facilita el proceso de implementación y contribuye a la reutilización de código.
 - Idea 1: Almacenar los datos en un hash table. Es una buena alternativa para almacenar los datos que se encarguen a través de un Excel, sin embargo, la implementación en código puede ser dispendiosa.
- Generar reporte por columnas:

Aproximaciones

- Idea 1: Se considera que la alternativa para generar el reporte por columnas es generar el reporte en una table view dado que es la única posible solución propuesta

- Generar gráficos estadísticos:

Aproximaciones:

- Idea 1: La idea de utilizar las librerías de .NET para generar gráficos de barras, pastel y dispersión se considera que es la mejor alternativa ya que es la que menos tiempo lleva poder realizar estas gráficas.
 - Idea 2: La idea de valerse de una herramienta de generación de gráficos brindada por un software tal como Excel es una buena alternativa, pero conllevaría mucho tiempo poder realizar las gráficas así.
- Clasificar la caracterización de riesgo de suicidio:

Aproximaciones:

- Idea 1: Para la caracterización de los riesgos de suicidio se vio como una buena opción optar por la alternativa de árboles de decisiones, pues este mecanismo no requiere la preparación de datos

excesivamente complejos y es válido tanto para variables cuantitativas como cualitativas, que son las que manejamos en nuestro sistema.

- Idea 2: árboles de regresión. Sin embargo, la idea implica un estudio más extensivo del concepto.
- Ambiente experimental:

Aproximaciones:

- Idea 1: Modelar en el sistema un módulo con herramientas básicas de análisis experimental como “Configuración de ambiente experimental, ejecución del ambiente y pruebas y reporte de resultados de las pruebas realizadas”.
- Idea 2: Incluir dentro del sistema a desarrollar una serie de métodos con la capacidad de interpretar las entradas de los tratamientos con los cuales se va a realizar el experimento y que estos mismos generen las salidas correspondientes al resultado de la variable estudiada y con esta información sacar las conclusiones pertinentes.
- Idea 1: El modelar en el sistema un módulo de herramientas básicas de análisis experimental es la mejor alternativa dada la naturaleza de los requerimientos
- Idea 2: Incluir dentro del sistema a desarrollar una serie de métodos con la capacidad de interpretar las entradas de los tratamientos con los cuales se va a realizar el experimento es buena alternativa, pero no la más eficiente.
- Análisis estadístico:

Aproximaciones:

- Idea 2: Se define como mejor alternativa hacer uso del software SPSS, el cual es especializado en análisis estadísticos y modelos preconstruidos, lo cual confirma un menor gasto de recursos en tiempo y esfuerzo.
- Idea 1: El uso de R Studio es una buena alternativa, sin embargo, no es una herramienta tan completa y fácil de usar como SPSS
- Informes:

Aproximaciones:

- Idea 2: Se realiza una recopilación de los análisis y reportes obtenidos y con estos generar un documento en Word donde se especifiquen las conclusiones de los resultados, se descartan las otras dos opciones ya que es más factible implementar, un algoritmo que nos den el informe en este tipo de documento y el cual puede ser hasta más práctico para el usuario que quiera compartir este documento.
- Idea 1: Excel es una buena idea para generar los reportes, sin embargo, al tener tantas herramientas de análisis puede llegar a ser un poco confusa.

FASE 5: EVALUACIÓN Y SELECCIÓN DE LA SOLUCIÓN

Se decide utilizar las siguientes herramientas:

	Idea 1	Idea 2	Idea 3	Resultado
Carga del dataset	5	2	NA	
Almacenar registro	3	NA	5	Idea 3
Generar reporte por columnas	5	3	NA	Idea 1
Generar gráficos estadísticos:	4	2	NA	Idea 1
Clasificar la caracterización de riesgo de suicidio	5	2	NA	Idea 1
Ambiente experimental	4	3	NA	Idea 1
Análisis estadístico	3	5	NA	Idea 2
Informes	3	5	NA	Idea 2

FASE 6: PREPARACIÓN DE REPORTES, PLANOS Y ESPECIFICACIONES

- [Diagrama de clases](#)
- [Diagrama de objetos](#)

FASE 7: IMPLEMENTACIÓN DEL DISEÑO

- [Repositorio en GitHub](#)