

Visualización de datos

Proyecto de fin de curso

Visualización de eventos catastróficos

1. Enunciado

1.1. Extracción de conocimiento a partir de datos

La extracción de conocimiento a partir de datos (KDD por sus siglas en inglés)¹ es un proceso que permite, a partir de un conjunto de datos, obtener conocimiento nuevo. Una de las últimas etapas de este proceso es la *interpretación y validación*, en la cual, uno de los objetivos es visualizar los resultados de la etapa de *data mining*.

En esta etapa de interpretación y validación, las herramientas de visualización juegan un papel importante para la culminación del proceso KDD. En efecto, la visualización permite presentar a los expertos del caso estudiado los hallazgos (información) encontrada en las etapas previas. En ese sentido, la visualización debe ser implementada de manera tal que los expertos en el tema permitan aprender de la información encontrada por los analistas de datos y puedan tomar decisiones correctas.

En este proyecto de fin de curso, aplicaremos el proceso KDD sobre un conjunto de datos, haciendo énfasis en la etapa de visualización.

1.2. Caso de estudio

Desde el inicio de los tiempos, las catástrofes producidas por la naturaleza han causado daños en las civilizaciones que han habitado nuestro planeta. Wikipedia dice que, “un desastre natural hace referencia a las enormes pérdidas materiales y vidas humanas ocasionadas por eventos o fenómenos naturales”. A lo largo de nuestra historia, hemos visto diferentes fenómenos, de los cuales, los tsunamis, erupciones volcánicas y temblores de tierra fueron los que más daños han producido.



Estos tres fenómenos han sido estudiados durante mucho tiempo a fin de comprender mejor cómo éstos se desarrollan y buscar una forma de estar preparados y evitar desenlaces trágicos. Una de las tareas principales, antes de estudiar estos fenómenos, es saber dónde ocurrieron y conocer ciertos detalles acerca del evento, como por ejemplo, la profundidad focal donde se originó un tsunami, la elevación de un volcán en erupción o el

¹U. Fayyad, G. P.-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. AI Magazine, 17(3):37-54, Fall 1996.

epicentro de un terremoto. Afortunadamente, el NOAA (National Oceanic and Atmospheric Administration NOAA, <https://www.ngdc.noaa.gov>) cuenta con una gran base de datos de distintos fenómenos, entre los cuales se encuentran datos relacionados a los tres fenómenos que estudiaremos en este trabajo de fin de curso:

1. Los tsunamis (<https://www.ngdc.noaa.gov/hazard/tsu.shtml>).
2. Las erupciones volcánicas (<https://www.ngdc.noaa.gov/hazard/volcano.shtml>).
3. Los temblores de tierra (<https://www.ngdc.noaa.gov/hazard/earthqk.shtml>)

2. Trabajo solicitado

2.1. Objetivo del proyecto

Este proyecto tiene por objetivo diseñar e implementar un prototipo de visualización para mostrar información interesante a partir de los tres juegos de datos asociados a catástrofes naturales.

/!\ Nota: Los datos a ser visualizados pueden variar en función de lo que los grupos de trabajo decidan. La elección de la temática es libre, sin embargo, los entregables son los mismos para todos los grupos.

2.2. Etapas que debe comprender el presente proyecto

En este proyecto seguiremos las etapas propuestas en el proceso KDD. Sin embargo, como se mencionó líneas arriba, nos concentraremos en la visualización.

1. **Pre-tratamiento de los datos:** utilice *Jupyter* y *python* para dar un formato correcto a los datos y almacenarlos en una estructura que permita su manipulación. En esta etapa, puede ser necesario agregar algunas columnas para geo-localizar entidades espaciales. Además, puede usar otros juegos de datos complementarios.
2. **Data Mining:** esta etapa puede ser obviada. Sin embargo, - si cree conveniente - puede usar alguna técnica básica de minería de datos (o EDA) a fin de generar algún otro indicador complementario a los que ya tienen en los datos originales.
3. **Visualización:** la visualización es la parte más importante de este trabajo. La idea es que utilicen una o más técnicas de visualización aprendidas en el curso para construir un prototipo que les permita alcanzar el objetivo propuesto en la Subsección anterior. Para este trabajo, las herramientas que podrían utilizar son *Python3*, *Jupyter*, *Plotly*, *Dash*, *Flask* u otro. El uso de *Dash* u otra librería podrá permitir que el prototipo sea accesible desde un navegador Web y que además permita a un usuario interactuar con él (por ejemplo, mostrar los datos agregados por continente, por país, etc.). Recuerden que están construyendo una herramienta para facilitar la toma de decisiones.

/!\ Nota: La calificación de este trabajo dependerá de cuán creativos sean y de las funcionalidades y bondades que ofrezca la propuesta al momento de ser utilizada por los usuarios finales.

3. Entregables

El trabajo final del curso tiene dos entregables:

1. [15 puntos] **El prototipo** completo (en Jupyter) comprimido en un archivo ZIP (u otro). No se olvide de documentar el código a fin de que sea comprensible.
2. [5 puntos] Un documento de cinco (5) páginas como máximo en el cual muestren la semántica y las opciones de la visualización y (un **mini manual**).

4. Grupos y fechas

Este trabajo deberá ser realizado en grupos de cinco (5) estudiantes como máximo. La **nota final** solamente estará compuesta por una nota grupal (el proyecto, sobre 20). No habrá una nota individual. La fecha de entrega de los trabajos se realizará el **sábado 30 de enero del 2021** indefectiblemente y todos los miembros del grupo deberán estar presentes. En esta fecha, cada grupo deberá presentar su proyecto al resto de compañeros en 15 minutos aproximadamente (+ 5 de preguntas)

Manos a los datos!!!