



Análisis exploratorio y Preprocesamiento de datos

Introducción

Parte importante del proceso de Minería de datos consiste en analizar el comportamiento inicial de los datos para determinar sesgos, restricciones y consideraciones en los datos que puedan influir en la toma de decisiones al final del flujo de trabajo. Para realizar estas exploraciones es necesario preprocesar los datos, eliminando o, en menor medida, limitando los problemas de formato y presentación de los mismos, así como presentar gráficos que soporten decisiones técnicas dentro y fuera del contexto del problema, como los tipos de modelos a usar, los tipos de tareas de preprocesamiento necesarias, consideraciones de las interpretaciones de los modelos resultantes, entre otros.

Para ejercitar dichas prácticas, se requiere que usted realice un curso interactivo en el lenguaje de programación R y documente los comandos y fragmentos de código que estén en dicho curso y considere de utilidad como parte del proceso de Minería de datos, que le sirva de caja de herramientas para ser usados en otros problemas, esto podrá ser entregado en el formato de su preferencia, tomando en cuenta que dicho formato deba facilitar el uso de los comandos y su documentación de los mismos. Se requiere que utilice la siguiente convención de nombre CI.PDF o CI1_CI2_CI3.PDF según requiera, por ejemplo: 24336448.PDF:

Consideraciones

- El presente trabajo puede ser realizado en grupos (de hasta 3 personas) o individualmente.
- El presente trabajo deberá ser entregado el día **16-09-19** a más tardar
- Se considerará la amplitud de los comandos documentados como aspecto merecedor de mayor puntaje, esto es, mientras mas documentas de cada tema documento podría recibir mayor puntaje.

Asignaciones:

Los cursos referidos a continuación se encuentran en swirl repo.

- Realice todo el curso de **Getting and Cleaning Data** documentando, al menos, los comandos que se explican en dicho curso y que resulten de utilidad en el proceso de minería de datos.
- Realice todo el curso de **Exploratory Data Analysis** documentando, al menos, los comandos que se explican en dicho curso y que resulten de utilidad en el proceso de minería de datos.



Ayuda:

Referencias de problemas a plantear:

- [Tutorial de swirl](#)
- [Refencia de cheatsheet](#) que pueden tomar como inspiración en cuanto a lo estético consideren que dicho documento les puede ser de utilidad fuera de la materia.

GDMD/I-2019