

Programa Ciencia de los Datos

Curso Minería de datos e Inteligencia de Negocios

Marzo 2020

Prof. Lorena Zúñiga

Proyecto

Valor 25%

Objetivos

- Plantear un problema de negocio para ser resuelto mediante técnicas de minería de datos.
- Aplicar parcialmente la metodología CRISP-DM para resolver un caso básico de minería de datos
- Aplicar principios de visualización de información para presentar mediante un dashboard los resultados de los modelos creados.

Descripción

El (la) estudiante planteará un problema de negocio (real o ficticio) que pueda resolverse mediante técnicas de machine learning y minería de datos, ya sean predictivas o descriptivas.

Tanto el problema como su solución inicial serán realizados siguiendo la metodología CRISP-DM vista en clase. Se documentarán las fases y actividades realizadas.

Para el desarrollo de las actividades que así lo requieran se utilizará el lenguaje de programación R o bien Python.

El dashboard deberá contener información relevante sobre los resultados de los modelos y se implementará con Tableau Desktop.

Fases y actividades a desarrollar:

- **Fase de Entendimiento del negocio**
 - Determinar los objetivos de negocio
 - Objetivos de negocio

- Criterios de éxito (en términos de negocio)
- Determinar los objetivos de minería de datos
 - Objetivos de minería de datos
 - Criterios de éxito desde la perspectiva de minería de datos
- **Fase de Entendimiento de los datos**
 - Descripción de los datos (descripción general de los datos a utilizar, cantidad de registros, tipos de datos, etc.)
 - Exploración de los datos
 - Verificación de la calidad de datos
- **Fase de Preparación de los datos**
 - Selección de los datos
 - Limpieza de los datos
 - Construcción de nuevos datos (atributos)
 - Transformaciones aplicadas a los datos

Si la preparación de los datos la realiza en un lenguaje de programación distinto a R, debe documentar los cambios o transformaciones aplicadas a los datos. Aporte screenshots y agréguelos como figuras en el documento RMarkdown.

- **Fase de modelado**
 - Selección de las técnicas a utilizar
 - Construcción del modelo
 - Selección de los parámetros (justificar selección)
 - Ejecución
 - Descripción de los modelos obtenidos
 - Evaluación de los modelos
 - Describir los resultados obtenidos con cada modelo
 - Seleccionar el mejor modelo

Diseño y construcción del dashboard

- Integrar los resultados obtenidos a partir del modelo seleccionado (ej: clusters, resultados de clasificación, reglas, etc) en un dashboard que facilite el análisis de dicha información.
- Puede integrar también los datos iniciales (después de transformarlos).
- Publique el dashboard en Tableau Public y agregue el enlace al documento RMarkdown.
- Exporte el documento RMarkdown a html

Formato de entrega: archivo html con la descripción y resultados de las diferentes fases CRISP-DM y el enlace al dashboard.

Forma de trabajo: individual o en parejas

Forma de entrega: subir documento html del proyecto al TECDigital

Fecha de entrega: hasta el lunes 06 de abril de 2020, a las 3:00 p.m.

Evaluación

Rubro	Valor
Fase de Entendimiento del negocio	10
Objetivos de negocio	3
Criterios de éxito (en términos de negocio)	2
Objetivos de minería de datos	3
Criterios de éxito (en términos de minería de datos)	2
Fase de Entendimiento de los datos	10
Descripción de los datos (descripción general de los datos a utilizar, cantidad de registros, tipos de datos, etc.)	5
Exploración de los datos	2.5
Verificación de la calidad de datos	2.5
Fase de Preparación de los datos	10
Selección de los datos	3
Limpieza de los datos	2
Construcción de nuevos datos	2
Transformaciones aplicadas	3
Fase de modelado	30
Selección de técnicas	5
Construcción de los modelos	10
Descripción de los modelos	5

Evaluación de los modelos y selección	10
Dashboard	30
Exposición	10
Total	100

- La selección de los grupos o personas que exponen su proyecto se hará al azar el mismo día de la entrega, durante la hora de clase.
- Se dispondrá de 15 a 20 minutos máximo para exponer el proyecto.
- Quienes no expongan deberán plantear al menos una pregunta a un grupo o persona que expone, de lo contrario no tendrán puntaje.