



ML2006 Visualización de Datos en Python

Descripción

Al hacer análisis de datos se invierten muchos recursos en la correcta preparación de los datos y modelado; no obstante, es primordial presentar los resultados de forma tal que sea muy rápido y fácil interpretarlos; como bien se dice, una buena imagen (gráfico) es más valiosa que mil palabras. De esta forma, en este curso se presentarán los fundamentos del lenguaje Python para visualizar la información por medio del uso de herramientas gráficas.

Adicionalmente se usarán paquetes que permiten visualizarde mejor forma los datos según su tipo y dominio. Es importante recordar que la visualización de la información no es únicamente para presentar resultados finales, sino también para conocer los datos sobre los cuales se plantean los modelos.

Objetivos

En este curso el estudiante será capaz de:

1. Comprender y aplicar correctamente los principios de visualización de datos.
2. Aplicar correctamente los principios de la teoría del color en el campo de la visualización de datos.
3. Seleccionar el gráfico correcto según los datos que se deban representar.
4. Diseñar gráficos estáticos utilizando las librerías matplotlib y seaborn.
5. Comprender y aplicar los principios de la teoría de la gramática de los gráficos.
6. Crear gráficos interactivos con Plotly.
7. Visualizar información geoespacial utilizando mapas estáticos e interactivos.
8. Diseñar reportes profesionales utilizando Jupyter Notebooks.
9. Diseñar aplicaciones web con Dash.

Metodología

Basado en la teoría y en la aplicación directa de los conceptos aprendidos. Para esto se dispondrán de las siguientes herramientas:

1. Una video conferencia semanal, las cuales quedarán grabadas en Zoom, para que los alumnos la puedan acceder en cualquier momento.
2. Trabajos prácticos semanales.
3. Foros para plantear dudas al tutor y compañeros.
4. Aula virtual en Moodle.

Luego de este curso el estudiante será capaz de:

Desarrollar proyectos de Minería de Datos que involucren aspectos de visualización de datos, además será capaz de desarrollar reportes o aplicaciones con gráficos estáticos o interactivos.

Contenido

1. **Fundamentos de la visualización de datos.**
 - a. Historia de la visualización de datos.
 - b. Principios de la codificación visual.
 - I. Forma.
 - II. Posición.
 - III. Color.
 - IV. Tamaño.
 - V. Transparencia.
 - VI. Grupos.
 - c. Principios de la teoría del color aplicados a la visualización de datos.
 - d. Teoría de la gramática de los datos.
 - e. Selección de tipos de gráficos según los datos a representar.
 - f. Buenas prácticas para la presentación de datos.
 - I. Orden de los datos.
 - II. Títulos y anotaciones.
 - III. Uso correcto de las tipografías.
 - IV. Selección de colores aptos para personas con daltonismo.



Programa Iberoamericano de
Formación en Minería de Datos

2. Visualización de datos geoespaciales.

- a. Sistemas de coordenadas.
- b. Manejo de datos geoespaciales con python.
- c. Mapas coropléticos.
- d. Mapas de calor o densidad.
- e. Análisis de las limitaciones del uso de divisiones político-administrativas para la visualización de datos.
- f. Gráficos alternativos para datos geoespaciales.
- g. Mapas interactivos con python.

3. Gráficos interactivos.

- a. Introducción a Plotly.
- b. Estructura básica de un gráfico interactivo en Plotly.
- c. Desarrollo de gráficos interactivos en Plotly.
- d. Configuración de gráficos interactivos en Plotly.

4. Creación de reportes.

- a. Creación de reportes con Jupyter Notebooks.
- b. Personalización de los reportes para adaptar al estilo de tu organización.
- c. Exportar reportes en distintos formatos.

5. Desarrollo de aplicaciones.

- a. Introducción a Dash.
- b. Layout.
- c. Interactividad.
- d. Introducción a Dash Bootstrap Components.

Evaluación

El curso se evalúa con 4 tareas, una por semana, cada tarea tiene un valor de 25 puntos. La nota mínima de aprobación es de 70.

Bibliografía

- [1] Abha Belorkar et al. Interactive Data Visualization with Python: Present your data as an effective and compelling story, 2nd Edition. Google-Books-ID: _1PdDwAAQBAJ. Packt Publishing Ltd, Apr. 14, 2020. 362 pp. isbn: 978-1-80020-106-4.
- [2] Ekaba Bisong. "Matplotlib and Seaborn". In: Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners. Ed. by Ekaba Bisong. Berkeley, CA: Apress, 2019. pp. 151-165. isbn: 978-1-4842-4470-8. doi: 10.1007/978-1-4842-4470-8_12. url: https://doi.org/10.1007/978-1-4842-4470-8_12 (visited on 10/21/2022).
- [3] Elias Dabbas. Interactive Dashboards and Data Apps with Plotly and Dash: Harness the power of a fully fledged frontend web framework in Python – no JavaScript required. Google-Books-ID: b_gqEAAQBAJ. Packt Publishing Ltd, May 21, 2021. 364 pp. isbn: 978-1-80056-035-2.
- [4] Sandro Tosi. Matplotlib for Python Developers. Google-Books-ID: 8Fs3AgAAQBAJ. Packt Publishing Ltd, Nov. 9, 2009. 525 pp. isbn: 978-1-84719-791-7.
- [5] Claus O. Wilke. Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. Google-Books-ID: XmmNDwAAQBAJ. "O'Reilly Media, Inc.", Mar. 18, 2019. 390 pp. isbn: 978-1-4920-3105-5.