

Taller 2: Analítica de datos 2022-1

Profesor: Andrés Marino Álvarez Meza, Ph.D.
Departamento de ingeniería eléctrica, electrónica y computación
Universidad Nacional de Colombia - sede Manizales

1. Instrucciones

- El taller debe ser enviado al correo electrónico `amalvarezme@unal.edu.co` desde su correo institucional (no se aceptarán envíos desde correos diferentes a `@unal.edu.co`) incluyendo las discusiones y desarrollos sobre celdas de texto en latex de Colaboratory y códigos en Python (celdas de código comentadas y discutidas sobre Colaboratory), referente a los ejercicios propuestos.
- Las secciones 2, 3 y 4 deben desarrollarse de forma individual. La sección 5 debe desarrollarse en los grupos del proyecto de curso (máximo hasta 3 personas). Respecto a la sección 5, enviar desde un solo correo los desarrollos, indicando los integrantes del grupo de trabajo.
- **Fecha máxima de entrega secciones 2, 3 y 4 (individual): 29 de junio de 2022.**
- **Fecha de sustentación sección 5 (exposición Dashboard/Presentación proyecto final -grupal): 29 de junio de 2022.**

2. Conceptos básicos en clasificación

- Realizar los ejercicios planteados en el cuaderno [Clasificadores Bayesianos](#).
- Realizar los ejercicios planteados en el cuaderno [Medidas de desempeño en clasificación](#).
- Realizar los ejercicios planteados en el cuaderno [Comparación clasificadores](#).

3. Aprendizaje no supervisado

- Discuta cuáles son las principales diferencias y aplicaciones relacionadas con aprendizaje supervisado y aprendizaje no supervisado (Ver [material de apoyo](#)).
- Realizar los ejercicios planteados en el cuaderno [Aprendizaje no supervisado - reducción de dimensión-conglomerados](#).

4. Aprendizaje profundo

- Discuta las principales diferencia entre los métodos clásicos de aprendizaje de máquina y los métodos de aprendizaje pro-

fundo (Ver [material de apoyo](#)). Tenga en cuenta los conceptos de optimización por gradiente descendiente con propagación hacia atrás y diferenciación automática (Ver [material de apoyo](#)).

- Realizar los ejercicios planteados en el cuaderno [Introducción al aprendizaje profundo](#).
- Realizar los ejercicios planteados en el cuaderno [API funcional y redes CNN](#).

5. Proyecto en analítica de datos

- Consulte qué son y para qué se utilizan los Dashboards en el ámbito empresarial. Presente una tabla comparativa con al menos tres herramientas para la construcción de Dashboards (paquetes de programación, plataformas, servicios en nube, etc.), indicando sus principales características técnicas, y sus ventajas y desventajas para su empleo en proyectos de analítica de datos.
- Según la base de datos escogida en el anteproyecto, presente una solución basada en analítica de datos mediante cuaderno de Colaboratory, Kaggle, y/o Dashboard (se tendrán en cuenta puntos de bonificación para quienes presenten la solución mediante Dashboard), que contemple, al menos: i) Análisis exploratorio básico con diagramas de pastel, caja, histogramas, y visualización 2D con técnicas de reducción de dimensión. ii) Predicción de la(s) salida(s) de interés (clasificación, regresión o conglomerados), probando al menos dos algoritmos clásicos y un modelo tipo aprendizaje profundo (tener en cuenta el estado del arte estudiado en el anteproyecto). iii) Gráficas que condensan la información extraída del mejor modelo, con las medidas de desempeño obtenidas (error cuadrático, absoluto, acierto, curvas roc, etc.) y la relevancia de los atributos procesados (interpretabilidad para explicar el por qué de la predicción).

Referencias

<https://github.com/amalvarezme/AnaliticaDatos>

Géron, A., (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems. O'Reilly Media.

Bishop, C. M. (2006). Pattern recognition. Machine learning, 128(9).