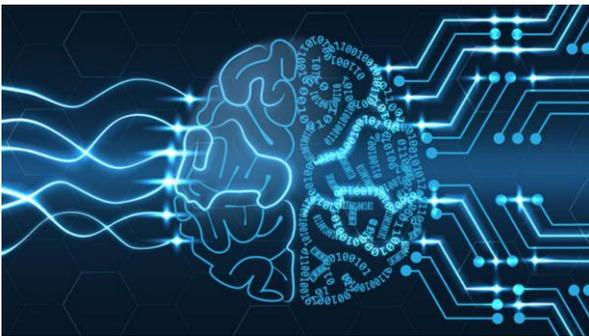


uah / Universidad
Alberto Hurtado

Universidad para **el bien común**



Fundamentos de Ciencia de Datos

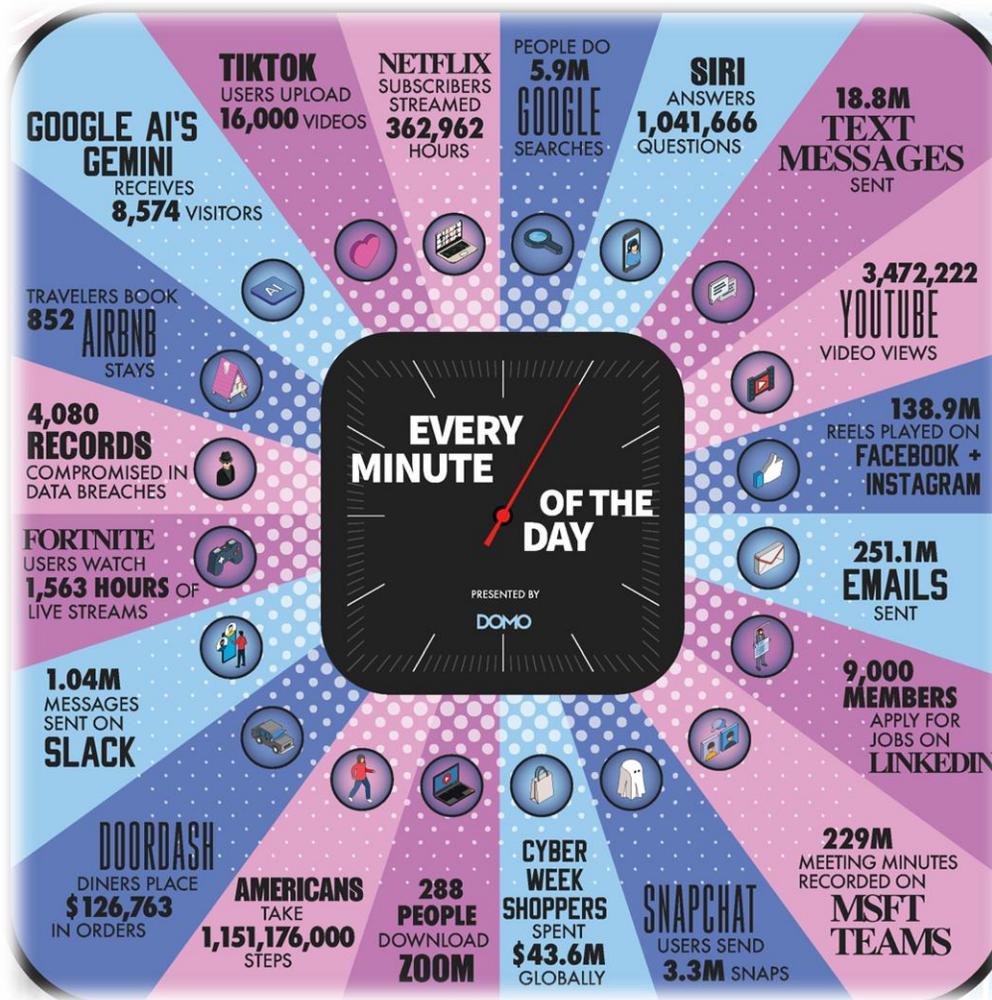
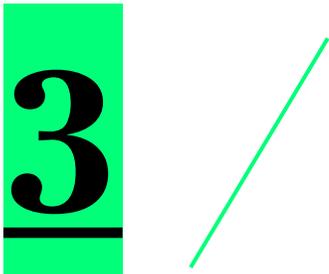
Facultad de Ingeniería
Wilson Alavia Medina, Académico
walavia@uahurtado.cl

2

Contenido

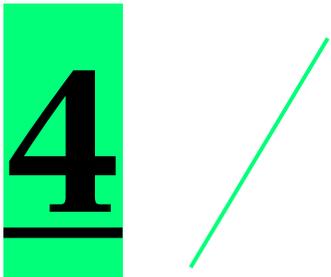
- Presentación profesor
- Motivación
- Descripción
- Propósitos de aprendizaje
- Contenido
- Metodología
- Evaluación de aprendizajes.
- Bibliografía

Motivación

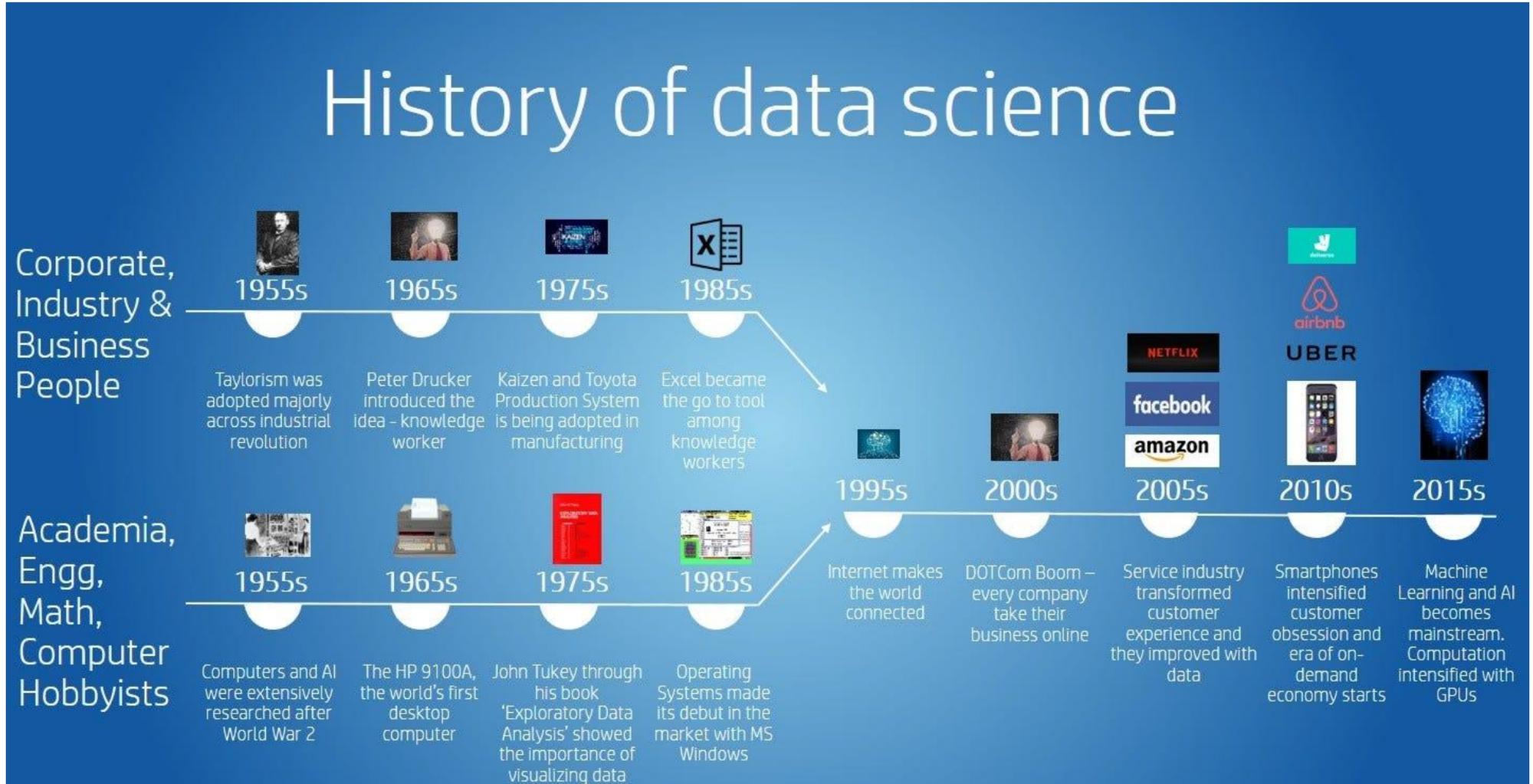


Fuente: Domo

Motivación

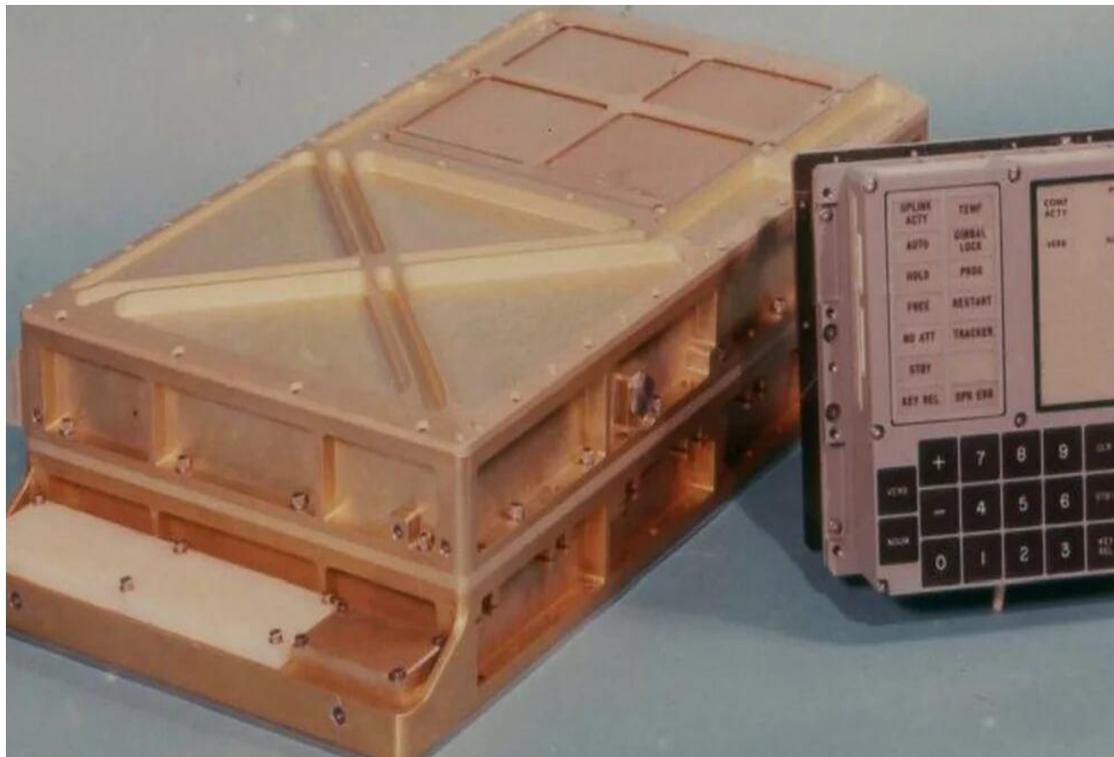


Motivación



Fuente: <https://datascienceacademy.ncsu.edu/2024/03/22/whm-ds/>
 Universidad para el bien común

6



Fuente: <https://www.infobae.com/america/50-anios-del-hombre-en-la-luna/2019/07/14/el-celular-que-llevas-en-el-bolsillo-es-mas-veloz-y-potente-que-la-computadora-que-se-uso-en-la-mision-apollo-11/>

63488 veces superior a la de aquel equipo y cuenta con una memoria RAM un millón de veces superior



Fuente: Wikimedia Commons

7

Margaret Hamilton



Ada Lovelace



Elizabeth Feinler



Florence Nightingale



Grace Hopper

Fuente: <https://datascienceacademy.ncsu.edu/2024/03/22/whm-ds/>

9



Fuente: <https://estebanairaldo.medium.com/tipos-de-lenguajes-de-programaci%C3%B3n-2be19e57d65a>

Profesor

10

Wilson Alavia Medina

Ingeniera Químico de la Universidad Técnica de Oruro, Magister en Ciencias de la Ingeniería mención Ingeniería de Procesos de Minerales y doctorado en Ingeniería de Procesos de Minerales de la Universidad de Antofagasta.

Áreas de trabajo: Desarrollo de procesos para minería metálica y no metálica, energías renovables y almacenamiento energético, modelación matemática y optimización, modelación multifísica, inteligencia artificial, comunicación segura, economía circular y sostenibilidad.

Ha trabajado en la UCN, UST, USACH y UTEM. Ha realizado investigación en el Max Planck Institute for Dynamics of Complex Technical Systems de Alemania.



11



Descripción

12

Esta actividad curricular, pertenece área de formación disciplinar y profesional (formación e iniciación a la disciplina y/o profesión), y se encuentra ubicada en el ciclo básico del plan de estudios. Busca desarrollar en el estudiantado capacidades para buscar soluciones a problemas y asistir la toma de decisiones a través de las metodologías básicas basadas en el uso de datos. Esto implica conocer la terminología básica, los conceptos técnicos, los desafíos tecnológicos de implementación y las implicancias éticas del uso de datos. Esta actividad curricular promueve un aprendizaje profundo de las estrategias fundamentales de la ciencia de datos sobre la base del autoaprendizaje, la autoevaluación, la autonomía y la autorregulación, dejando en segundo plano las herramientas; todo esto debido a la importancia de las primeras y la naturaleza cambiante de las últimas

Propósito de Aprendizaje

13

Al término de esta actividad curricular, el estudiantado será capaz de:

1. Analizar de forma crítica cuándo y cómo utilizar soluciones basadas en datos en problemas reales, considerando las oportunidades y dilemas éticos del uso de estas técnicas.
2. Implementar soluciones efectivas de principio a fin a problemas utilizando las estrategias más básicas basadas en datos.
3. Juzgar de forma crítica las limitaciones y desafíos pendientes del uso de este tipo de metodologías basadas en datos.
4. Construir plataformas para interactuar, refinar y visualizar los resultados de los métodos básicos basados en la ciencia de datos.

Contenido

14

Unidad I: Ciencia de Datos y Big Data

Unidad II: Temas éticos en la ciencia de datos

Unidad III: Estadística descriptiva

Unidad IV: Modelos predictivos – Regresión

Unidad V: Espacios de características

Unidad VI: Modelos discriminativos y generativos – Problema de clasificación

Unidad VII: Visualización de Datos

Contenido

15

Unidad I: Ciencia de Datos y Big Data

1. Qué es Big Data: definiciones, desafíos e infraestructura
2. Qué es ciencia de datos: visión de la disciplina, tareas, labores y toma de decisiones
3. Ingeniería de datos, minería de datos, analista de datos, ciencia de datos
4. Modelos predictivos, discriminativos y generativos
5. Modelos supervisados y no-supervisados
6. Repaso de distribuciones, media y varianza muestral, normalidad
7. Principios básicos de estadística bayesiana

Contenido

16

Unidad II: Temas éticos en la ciencia de datos

1. Sesgos y diversidad
2. Transparencia e interpretabilidad: hacia algoritmos justos
3. Privacidad y protección: límites de la anonimización

Unidad III: Estadística descriptiva

1. Actividad: desafío de reporte y descripción de datos reales
2. Tipos, formatos, fuentes y métodos de extracción de datos
3. Exploración inicial de datos
4. Preparación de los datos – tratamiento, limpieza e integración de datos de distintas fuentes
5. Construcción de nuevas variables e índices

Contenido

17

Unidad IV: Modelos predictivos – Regresión

1. Actividad: desafío de predicción con alguna base de datos pública
2. Mínimos cuadrados
3. Formulación bayesiana conexión con problemas inversos
4. Interpretación de los resultados

Unidad V: Espacios de características

1. Generación de características de los datos y concatenación
2. Normalización
3. Selección de características
4. Reducción de dimensionalidad (PCA)

Contenido

18

Unidad VI: Modelos discriminativos y generativos – Problema de clasificación

1. Actividad: desafío de clasificación
2. Modelos supervisados
3. Modelos no-supervisados
4. Métricas de rendimiento – Curvas ROC, AUOC, matriz de confusión
5. Balance de las clases y técnicas de mitigación de desbalance

Unidad VII: Visualización de Datos

1. Modelos y herramientas para visualización de datos de distinta naturaleza
2. Diseño y construcción de interfaces de reporte de indicadores de desempeño
3. Desafío final: construcción sistema end-to-end

Evaluación de Aprendizajes

19

- Informes escritos
- Proyecto aplicado/diseñado
- Presentaciones orales

Evaluación de Aprendizajes

20

- Primer parcial, 9 de Abril de 2024. Ponderación: 25 %
- Segundo parcial, 4 de Junio de 2024. Ponderación: 25 %
- Taller 1, 23 de Abril de 2024. Ponderación: 10 %
- Taller 2, 18 de Junio de 2024. Ponderación: 10 %
- Examen, 2 de Julio de 2024. Ponderación: 30 %

El requisito mínimo de asistencia a clases es del 50 %. La nota mínima de aprobación es de 4,0.

Cualquier conducta de fraude académico en las actividades de evaluación será calificada con nota 1,0. Si existe un comportamiento reiterado en ese sentido, este será informado a la Dirección de carrera para determinar la sanción respectiva.

Evaluación de Aprendizajes

21

- Los requisitos para eximirse del examen son los siguientes:
 - Asistencia igual o mayor al 70% de las clases.
 - Promedio de nota de evaluaciones parciales igual o mayor a 5,5.
 - Promedio de nota de talleres igual o mayor a 5,5.
- En caso de inasistencia justificada a una prueba parcial el examen sustituye esa nota.
- La nota de examen no sustituye una calificación de prueba parcial si esta corresponde a una sanción por fraude académico.
- Los justificativos de inasistencia se deben entregar a la Coordinadora Académica de la Facultad de Ingeniería.

Bibliografía

22

- Bibliografía básica
 1. Joel Grus. (2019). Data Science from Scratch: First Principles with Python. Sebastopol, CA, Estado Unidos, Editorial O'Reilly.
 2. Wes McKinney (2018). Python for Data Analysis: Data wrangling with Pandas, Numpy and IPython. Sebastopol, CA, Estados Unidos, Editorial O'Reilly.
 3. Bobadilla, J. (2020). Machine Learning y Deep Learning: Usando Python, Scikit y Keras. 1. Paracuellos de Jarama, Madrid, RA-MA Editorial.

Bibliografía

23

- Bibliografía complementaria

1. Menoyo Ros, D. García López, E. y García Cabot, A. (2021). Fundamentos de la ciencia de datos. Alcalá de Henares, España, Editorial Universidad de Alcalá.

2. Vicente Vírseda, J. A. González Arias, J. y Parra Rodríguez, F. J. (2019). Métodos de Data Science aplicados a la Economía y a la Dirección y Administración de Empresas. Madrid, UNED - Universidad Nacional de Educación a Distancia.

Muchas gracias.



uah/ Universidad
Alberto Hurtado