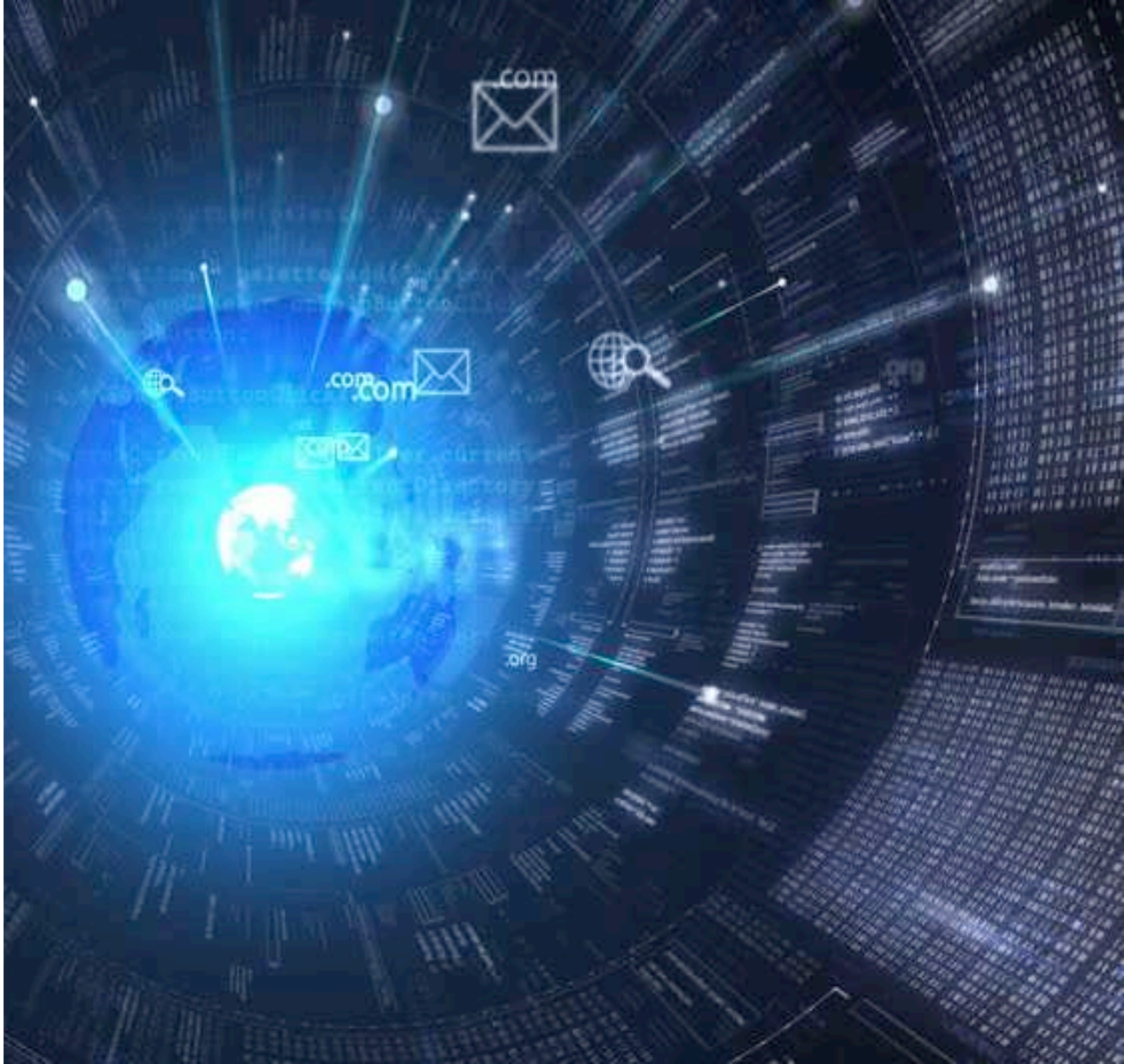
BIENVENIDOS A

DigitalHouse >
Coding School

Data Science



EQUIPO DE DATA SCIENCE

GERMÁN

COORDINADOR

MARTÍN

PRODUCTOR

PAOLO

PROFESOR

JULIÁN

AYUDANTE

PROFESORES

EQUIPO DE DATA

María Frances Gaska

Programadora orientada al Data Mining. Licenciada en Economía Magna Cum Laude por la Universidad de Buenos Aires. Especializada en Explotación de Datos y Descubrimiento del Conocimiento en la Facultad de Ciencias Exactas.

Germán Rosati

Sociólogo, Doctor en Ciencias Sociales (UBA) y Master en Generación y Análisis de Información Estadística (UNTREF). Especializado en la aplicación de métodos de machine learning/ data mining a las ciencias sociales.

Paolo Donizetti

Lic. Economía (UBA), Lic.Matemáticas (Roma - La Sapienza), Máster en Investigación de Mercado y Data Mining (U. di Bologna). Trabajó en Coca-Cola y el Grupo Techint. Hoy es productor audiovisual y se especializa en Machine Learning aplicado a gestión de contenidos.

Marianela Sarabia

Investigadora en desarrollo económico y consultora independiente.
Lic. en Economía (UBA), MSc en Economía Laboral Aplicada (SciencesPo y U.Turín) y doctoranda en Economía (UADE)

EQUIPO DE DATA

Leonardo Córdoba

Especialista en Data Mining y Economista de la UBA. Con experiencia como científico de datos en el sector privado y en el sector público, también me he desempeñado en la academia.

Martín Ríos

Estudiante avanzado de Ingeniería Civil. Especializado en Data Science también cuenta con experiencia en desarrollo y diseño Web.

Pablo Roccatagliata

Maestría en Estadística Matemática (en curso). .Posgrado en Economía. Universidad Torcuato Di Tella. de maestría. Licenciado en Economía. Universidad de Buenos Aires. Magna Cum Laude.

Demian Avendaño

Estudiante de Licenciatura de Ciencias Biológicas con orientación en Bioinformática, cursando materias de Licenciatura en Ciencias de la Computación.

EQUIPO DE DATA

Pablo Lorenzatto

Ingeniero de Datos y Machine Learning y Data Science.

Egresado de Ingeniería en Sistemas (UTN) especializado en Data Mining (UBA).

Amplia experiencia en la industria de datos.

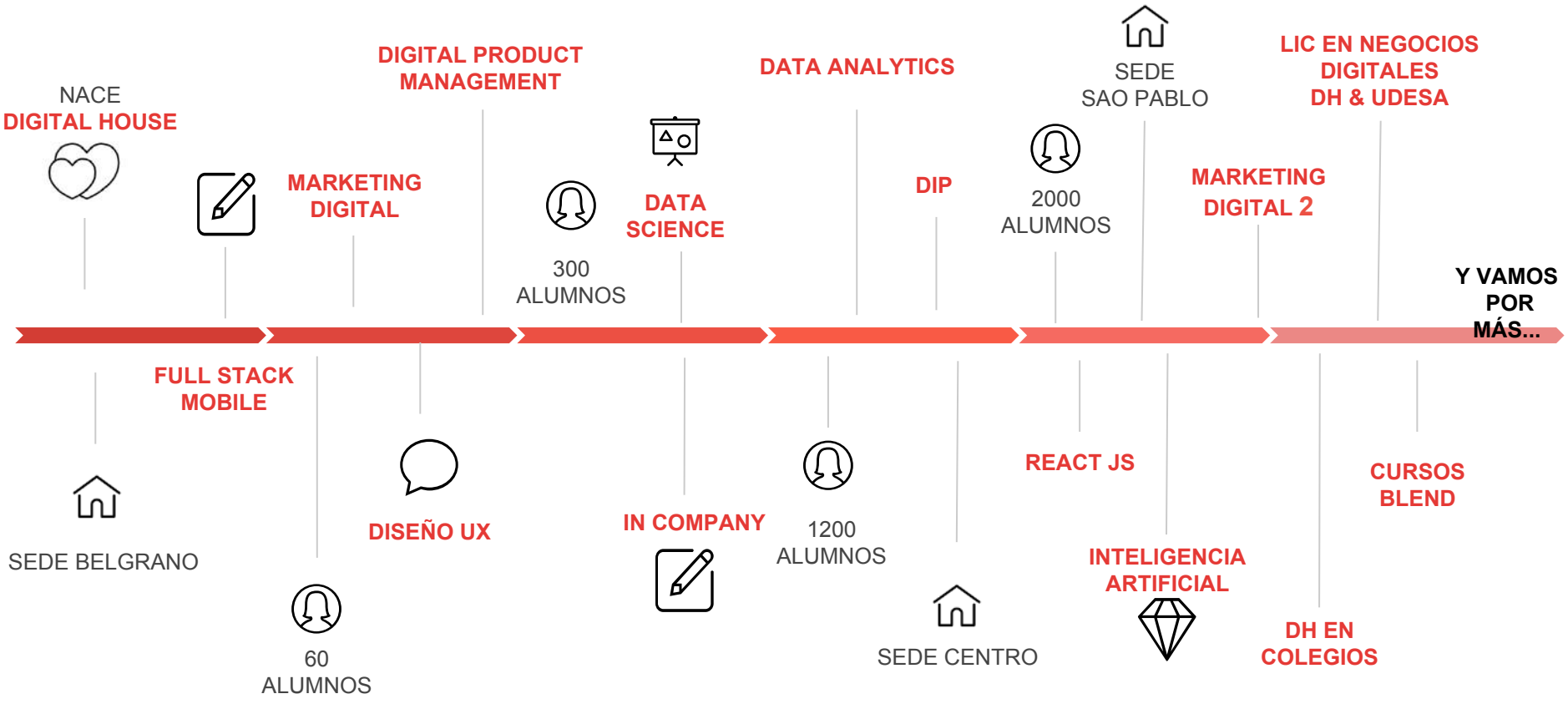
Matías Grinberg

Estudiante de Neurociencias (U. Favaloro). Colaborador en laboratorios de investigación y Data Scientist en la Industria. En camino hacia el área de Neurociencias Computacionales.

Julián Ansaldo

Licenciado en Economía (UBA). Trabajó como analista de datos en el Gobierno de la Ciudad de Buenos Aires en el marco del proyecto de integración del Barrio 31. Se desempeñó como ayudante docente de Economía Internacional y Teoría Política.

SOBRE NOSOTROS



¿Qué es el co learning?

Es un espacio para:

- Aprender junto a otros y trabajar en equipo.
- Hacer consultas puntuales de temas desarrollados durante las clases.
- Pensar y concretar nuevos proyectos.

Horario del espacio de co-learning:
Lunes a Viernes: 8:30hs a 22hs

Horarios de consulta a profesores:
Coordinar por Slack

¿Qué no es el co learning?

- No es un espacio de clases particulares, ni de consultoría.

- En el co learning no se recuperan las clases en las que el alumno estuvo ausente.

- En caso de ausencia, el alumno puede acceder al material en el Campus Virtual, y asistir al co learning para aclarar dudas puntuales.

DEPARTAMENTO DE ALUMNOS

Ángeles Castagnino

Juan Manuel Cestari

Gonzalo Galizia

Malena Schliserman

Horario de atención: 9.00 a 20.00 hs.

- Acompañamiento durante la cursada
- Soporte para acceso a campus virtual
- Asesoramiento sobre beneficios de Comunidad Digital House

¡Nuestras acciones comienzan a **mitad de semestre**!

Isabel Fortuna

Natalia Acosta

- Entrevistas de **Coaching** personalizado
- **Talleres** con estrategias para Cv Tech, **Linkedin**, Comunicación no Verbal en Entrevistas.
- Estrategias a **freelancers** y **emprendedores**.

María Eugenia Gallay

¡Unite a nuestros grupos!

Comunidad DH

Comunidad DH

Novedades - Eventos - Beneficios

BOLSA DE TRABAJO

¡Conocela!

DEPARTAMENTO DE PEDAGOGÍA

Johanna Litvinoff

Pablo Dalessandria

Daniela Sokn

Carla Arias

Julián Sandoval

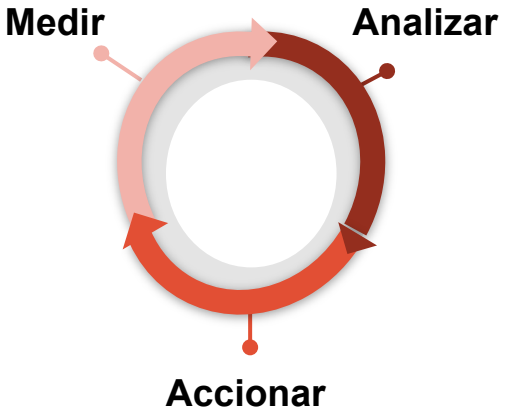
- Acompañamiento didáctico-pedagógico
- Formación continua del equipo docente
- Observación de clases
- Análisis de encuestas

ADMINISTRACIÓN Y COBRANZAS

Horario de atención: Lunes a Jueves de 9.00 a 20.00 hs.
Viernes de 9:00 a 18:00 hs.

Asesoramiento sobre pago de cuotas, vencimientos, medios de pago.

Nos interesa tu opinión!



¿Para qué?

- Mejora constante
- Cambios de acuerdo a las necesidades del grupo
- Reforzar las buenas prácticas

- Obligatorias
- Cada tres semanas
- Link vía mail
- Tenés 4 días para responderlas

Eventos para alumnos

DH SUMMIT

RECRUITING DAY

OPEN LAB

¿Están listos?

¿CÓMO APRUEBO EL CURSO?

CRITERIOS DE APROBACIÓN

90% DE ASISTENCIA

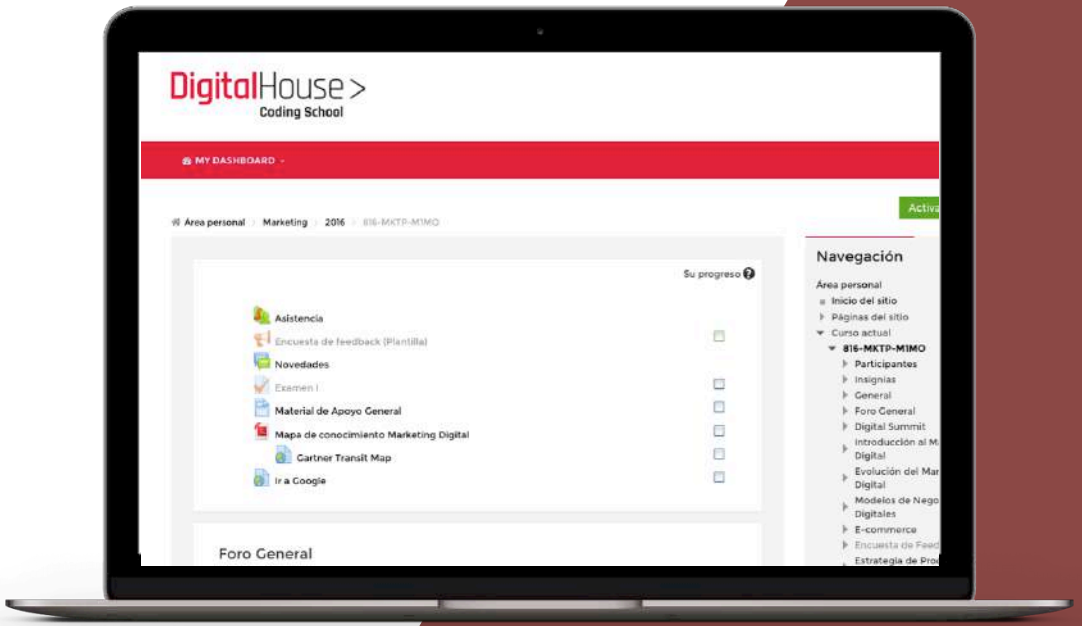
APROBACIÓN DE TRABAJOS PARCIALES

APROBACIÓN DEL PROYECTO INTEGRADOR

¡ESTEMOS CONECTADOS!

Qué canales vamos a usar y con qué finalidad cada uno

¿CÓMO NOS COMUNICAMOS?



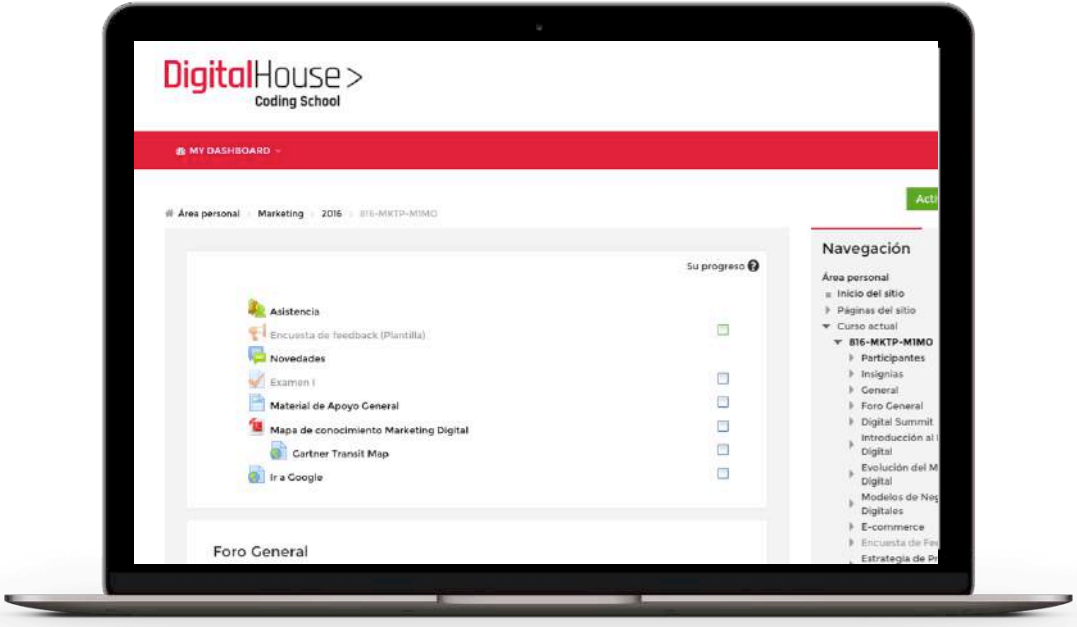
CAMPUS VIRTUAL

¿PARA QUÉ SIRVE?

- VER MATERIAL DE CLASES
- ENTERARSE DE EVENTOS, CHARLAS, ETC
- FORO: CONSULTAS O COMUNICADOS
- ENTREGA FORMAL DE TRABAJOS

WIFI: alumnos-digitalhouse

CLAVE: alumnos-digitalhouse



1 **INGRESAR EN:**

http://campus.digitalhouse.com

2 **DARSE DE ALTA CON:**

EMAIL:

CLAVE: Digitalhouse860!

3 **SUBIR FOTO! :)**

Ante cualquier inconveniente, enviar un mail a campus@digitalhouse.com

Aclarar:

Nombre y apellido.
Curso y comisión.
Describir brevemente cuál es el problema.

SLACK

dsdh-curso.slack.com

Para estar comunicados entre nosotros, compartir novedades e información relevante

MAIL

pdonizetti@digitalhouse.com

jansaldo@digitalhouse.com

Para notificar a los profesores cuestiones importantes o de urgencia

VAMOS A HACER UN CONTRATO

90% DE ASISTENCIA

PARTICIPACIÓN EN CLASE

TRABAJO EN PROYECTO INTEGRADOR DURANTE TODA LA CURSADA

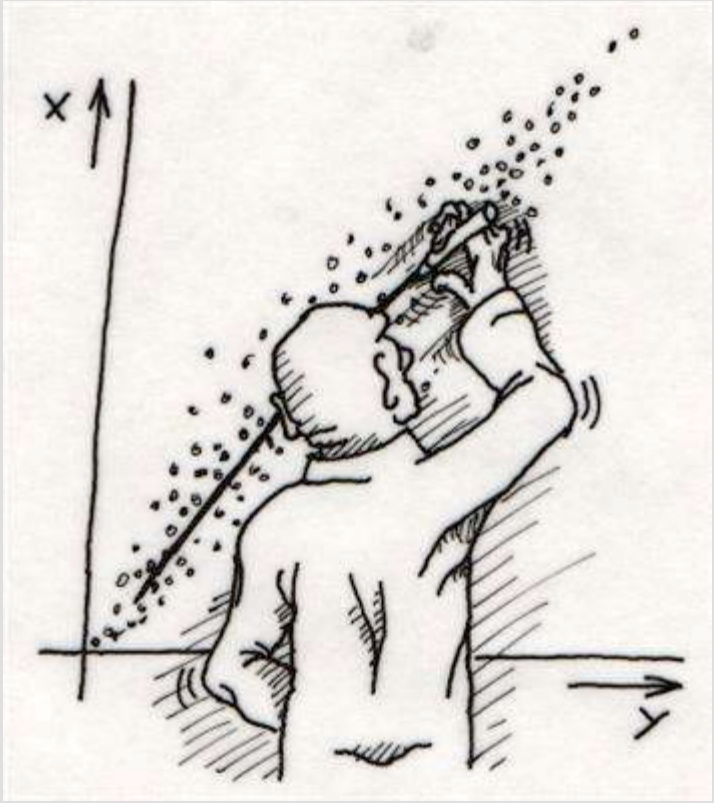
ENTREGA A TIEMPO DE LOS TRABAJOS

CUIDAR LOS ESPACIOS Y RECURSOS DE TRABAJO

APROVECHAR EL CO LEARNING

GENERAR UN BUEN CLIMA DE TRABAJO

¡COMENCEMOS!



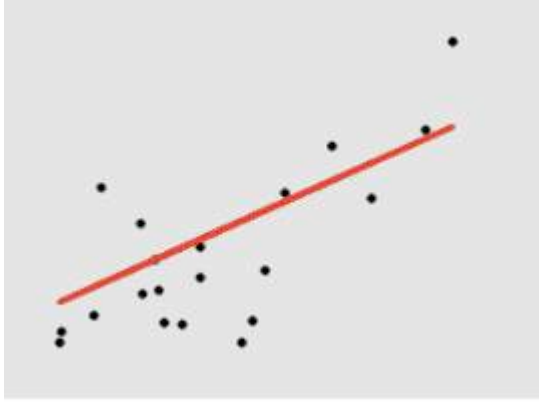
DigitalHouse > Coding School

DATA SCIENCE

UNIDAD 1
MÓDULO 1

Presentación del programa

1. **Presentar la filosofía y los objetivos del programa de Data Science**
2. **Desarrollar lineamientos de clase**
3. **Discutir sobre la naturaleza de la Ciencia de Datos**
4. **Lograr que los participantes del programa se presenten y se conozcan usando el flujo de trabajo de Data Science**



anokas

Mikel Bober-Irizar

Guildford, England, United Kingdom

Joined 3 years ago · last seen in the past day

https://mxbi.net

Followers 1386

Following 41

Competitions Master

With his skill, **enthusiasm**, and **cooperative attitude** within the community, Mikel is very much the template of a **rising star** in the Kaggle and greater AI communities.

Except for one thing: Mikel is just 16 years old.

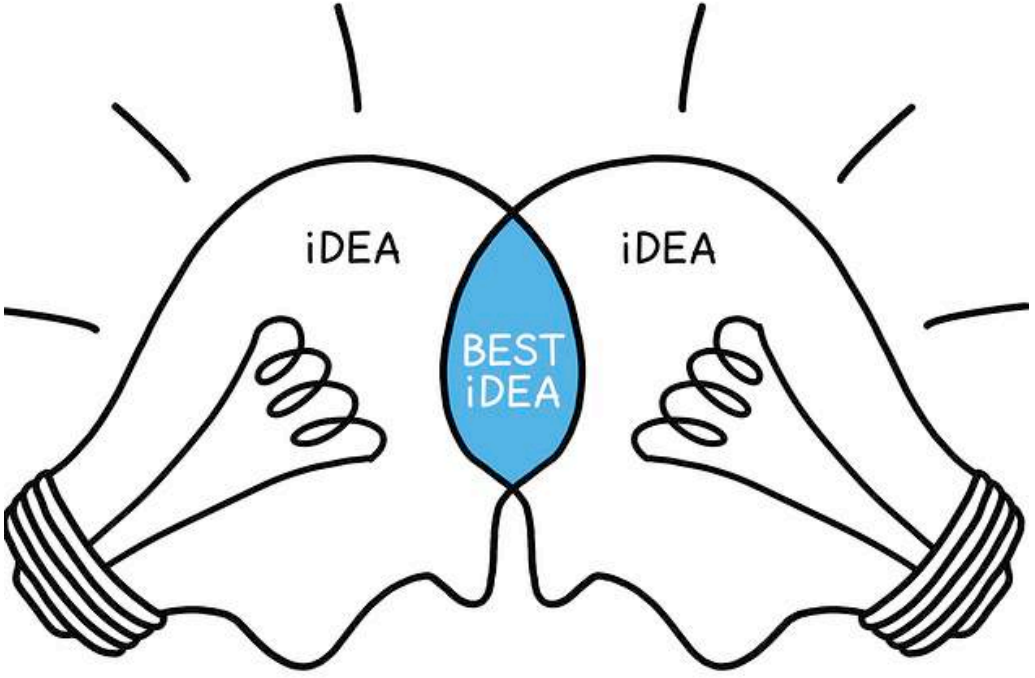
"I don't know all the math behind the algorithms, but in terms of actually using it, I think it's much more important to have a logical understanding of how it works. Even if I can't write it from scratch, I still know what it does, and that helps me to understand where it might be useful."

1. **Aprender las bases**
2. **Aprender a pensar**
3. **Aprender haciendo**
4. **Aprender a aprender**

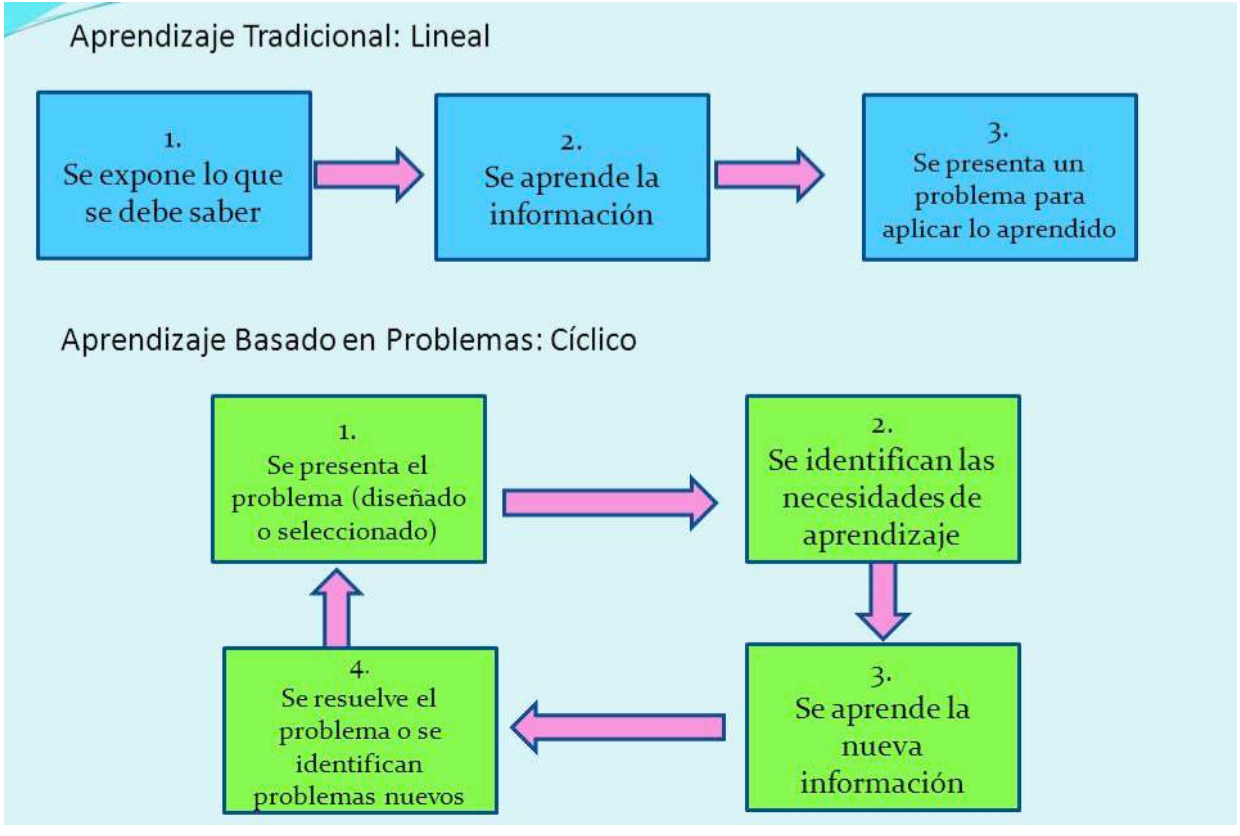
FILOSOFIA DEL PROGRAMA

- Fomentar y trabajar en un **entorno diverso**
- Encontrar el **ritmo de aprendizaje óptimo** para cada uno
- **Comunicar** pronto y frecuentemente
- El **éxito** en este curso no se obtiene por comparación. “There is nothing noble in being superior to your fellow man; true nobility is being superior to your former self.” Ernest Hemingway.

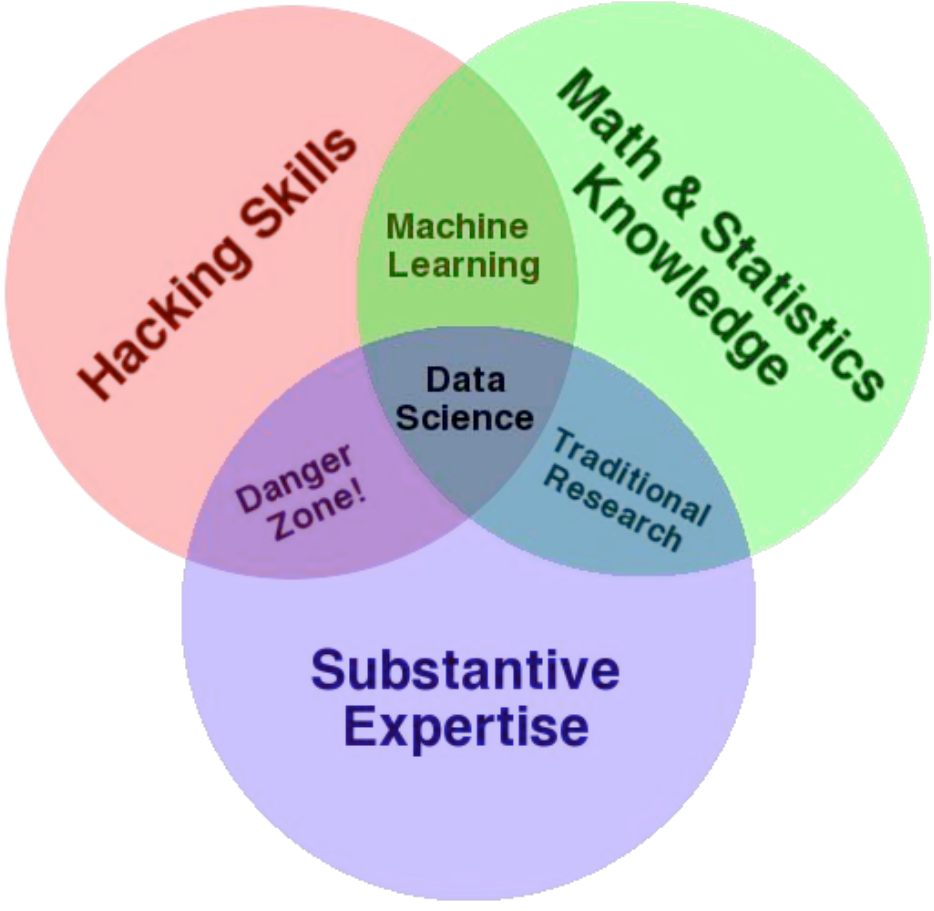
- La **dedicación**, más importante que el conocimiento previo
- Hacé **preguntas**, todo el tiempo, por default
- **Ayudá** a tus compañeros
- Sé **paciente** con vos mismo



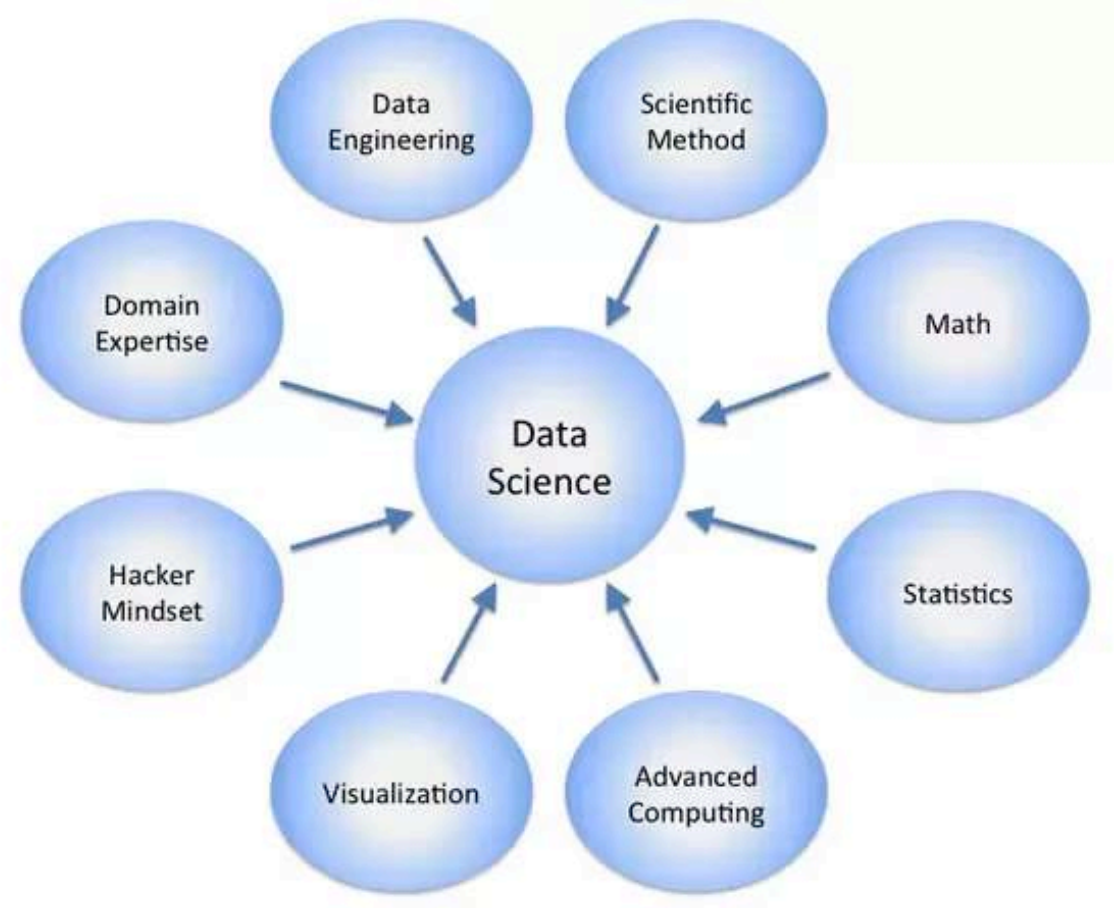
APRENDIZAJE BASADO EN PROBLEMAS



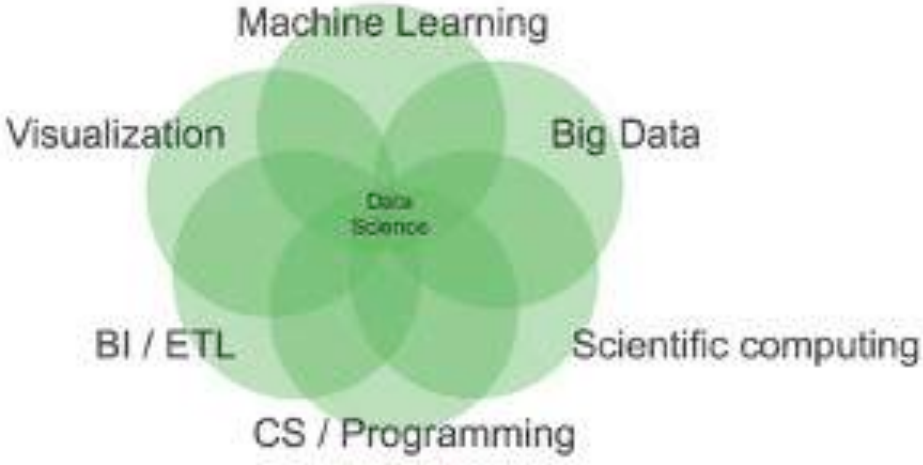
¿QUÉ ES DATA SCIENCE?



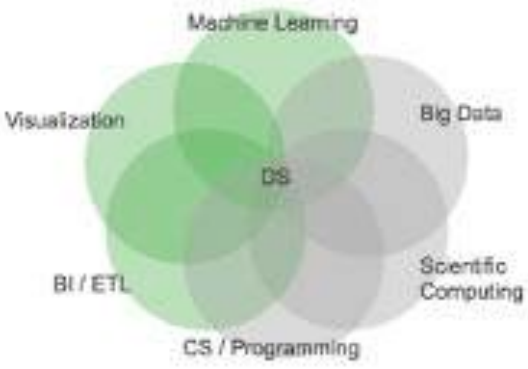
- Un set de herramientas y técnicas para extraer información útil de los datos
- Una práctica interdisciplinaria orientada a **resolver problemas**
- La aplicación de técnicas científicas a problemas prácticos
- ¿Quién usa Data Science?
  - Recomendaciones de películas Netflix
  - Algoritmo Amazon: “si te gustó X, quizás te guste Y”
  - Five Thirty Eight: cobertura electoral y de deportes
  - Google: traductores automáticos y sugerencias de búsquedas
  - Clasificadores de textos, imágenes, etc.


Data Engineering
Scientific Method
Domain Expertise
Math
Data Science
Hacker Mindset
Statistics
Visualization
Advanced Computing

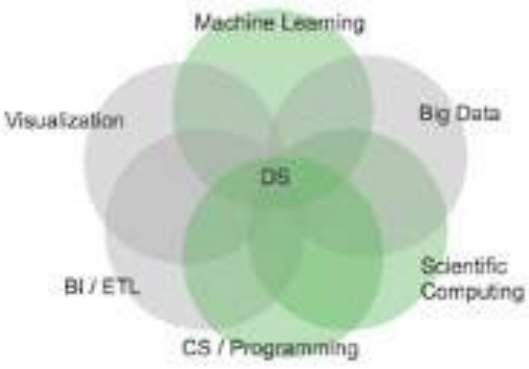
Traditional Data Science Venn Diagram



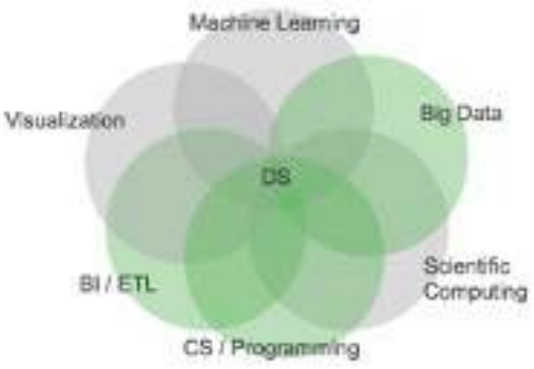
Revisited Data Science Venn Diagram



Statistician / Analyst



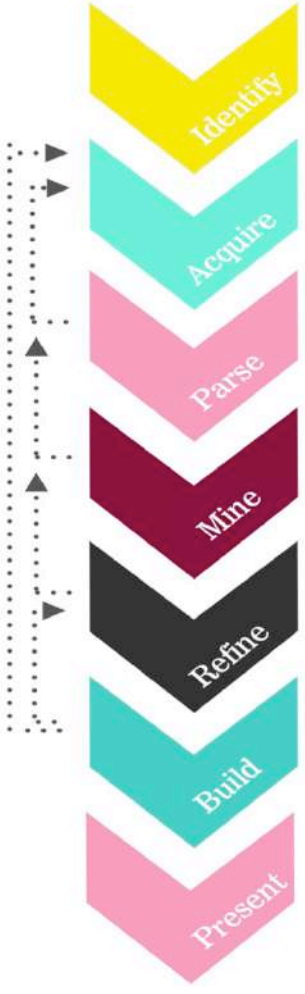
Research / Computational Scientist



Developer / Engineer

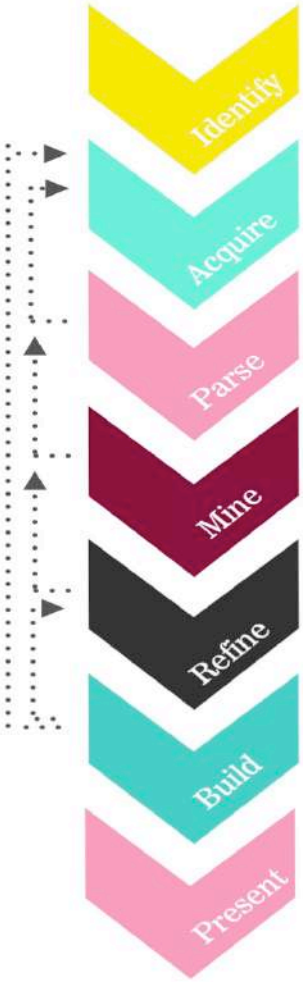
- A lo largo de las clases seguiremos el “Flujo de trabajo de Data Science”. Nos servirá para generar resultados confiables y reproducibles.
  - “confiables” = precisos
  - “reproducibles” = otros pueden replicar lo realizado y obtener resultados similares

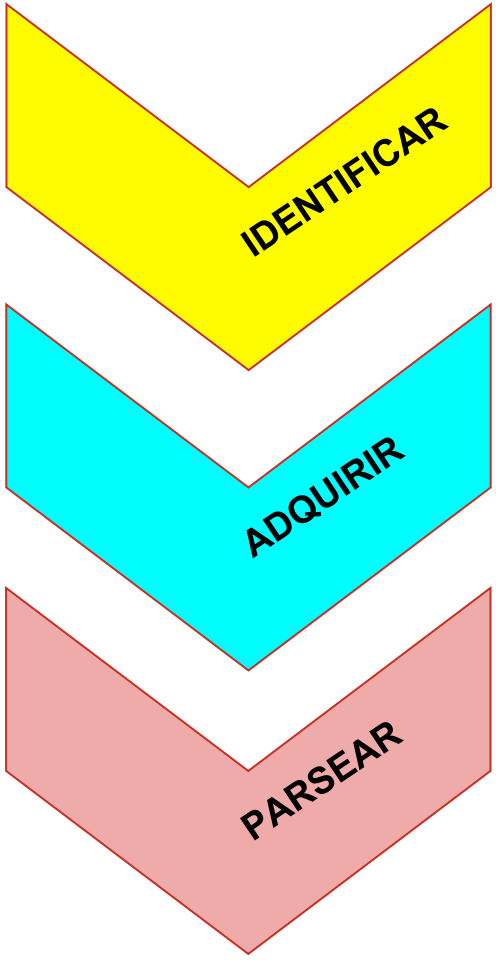
- En cualquier punto del proceso, puede ser necesario repetir pasos previos para iterar a lo largo del flujo. Esto dependerá de
  - la aparición de nuevos datos,
  - la necesidad de corregir errores,
  - el cambio acerca de las preguntas y objetivos, etc.



- El “Flujo de trabajo de Data Science” constituye, en última instancia, un set de standards sumamente útil y una referencia para tener en cuenta en los **desafíos del curso**.

- Repasemos las diferentes etapas, que están explicadas en detalle en el documento “**Flujo de Trabajo en Data Science.pdf**”





IDENTIFICAR EL PROBLEMA

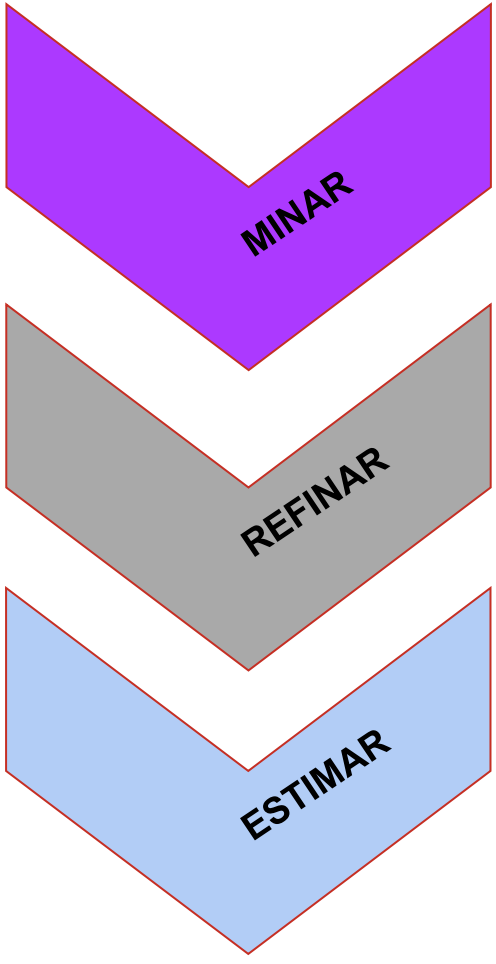
- Identificar los objetivos del producto/negocio/problema
- Identificar y generar hipótesis sobre metas y criterios para el éxito del análisis
- Generar un set de preguntas para identificar el dataset "correcto".

ADQUIRIR LOS DATOS

- Identificar el dataset "correcto"
- Importar los datos y generar las estructuras de datos adecuadas
- Determinar las herramientas más apropiadas para trabajar con los datos

PARSEAR LOS DATOS

- Explorar toda la documentación relacionada con los datos
- Realizar Análisis Exploratorio de los Datos (AED)
- Verificar la calidad de los datos



MINAR LOS DATOS

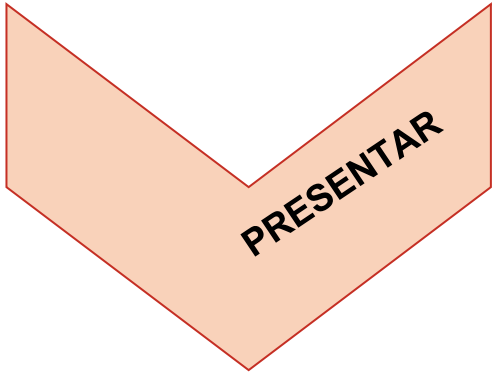
- Dar formato, limpiar, homogeneizar y filtrar los datos
- Crear nuevas columnas derivadas de los datos originales (recodificaciones, cálculos, etc.)

REFINAR LOS DATOS

- Identificar tendencias y outliers
- Aplicar y calcular estadísticos descriptivos e inferenciales
- Documentar y transformar los datos

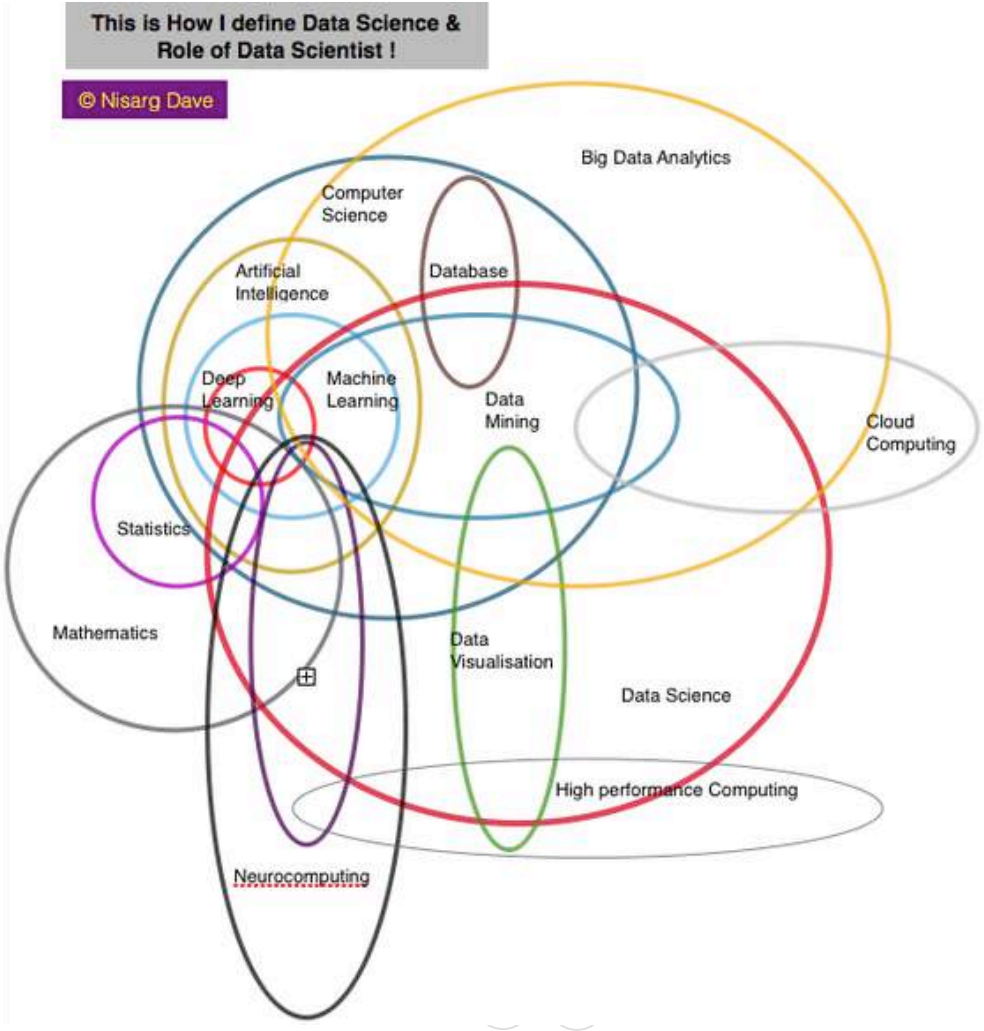
ESTIMAR UN MODELO

- Seleccionar un modelo apropiado (forma funcional, estimación, etc.)
- Estimar el modelo
- Evaluar y refinar el modelo



PRESENTAR LOS RESULTADOS

- Resumir los resultados del análisis con alguna narrativa o historia
- Presentar las limitaciones, los supuestos y las fortalezas del/ los modelo/s estimados
- Identificar preguntas derivadas y nuevos problemas para seguir profundizando el análisis


This is How I define Data Science & Role of Data Scientist !
© Nisarg Dave
Big Data Analytics
Computer Science
Artificial Intelligence
Database
Deep Learning
Machine Learning
Data Mining
Cloud Computing
Statistics
Mathematics
Data Visualisation
Data Science
High performance Computing
Neurocomputing


FAIL

MÓDULOS

- Que los asistentes sean capaces de

    - Extraer, consultar, limpiar y agregar datos para su análisis.
    - Construir, implementar y evaluar problemas de Data Science usando los algoritmos apropiados de machine learning.
    - Usar las herramientas de visualización adecuadas para comunicar sus conclusiones.
    - Investigar, modelar y validar procesos de resolución de problemas aplicados a datasets provenientes de diversas industrias para proveer experiencias en distintos tipos de problemas y soluciones del mundo real.

PROGRAMA - MÓDULOS

01 Fundamentos: Numpy, SQL y Visualización

02 EDA, Pandas & Limpieza de datos, Inferencia Estadística

03 Intro a ML: Regresión Lineal, Regularización, Validación de Modelos ,API de Sklearn, Web Scrapping

04 Problemas de Clasificación, GridSearch, Pipelines y Procesamiento distribuido

05 Text Mining, Series de Tiempo

06 Árboles y Métodos de Ensamble

07 Aprendizaje no supervisado: PCA, Clustering, Manifold

PROYECTO INTEGRADOR

Fundamentos: Numpy, SQL y Visualización

01

- Introducción al programa y a la disciplina
- Repaso de Python
- Estadística Descriptiva con Numpy
- Introducción a la Visualización de Datos
- SQL

EDA, Pandas & Limpieza de datos

02

- Pandas
- Probabilidad
- Limpieza de Datos
- Variables Dummies
- Datos Faltantes
- Estadística Inferencial
- Joins con Pandas
- Visualización

Desafío del Módulo

Usando un dataset crudo de Properati usarán Pandas para limpiar los datos, plantearán formalmente un problema y realizarán análisis exploratorio

Intro a ML: Regresión Lineal, Regularización, Validación de Modelos ,API de Sklearn, Web Scrapping

03

- Introducción a Machine Learning
- Regresión Lineal
- Intro a Stats Models & Sklearn
- Descomposición Bias-Varianza
- Regularización & Sobreajuste (Overfitting)
- Separación Entrenamiento/Test
- Web Scrapping
- Métricas de Regresión & Funciones de Pérdida (Loss Functions)
- Descenso del gradiente
- Feature Scaling (Normalización)

Desafío del Módulo

Los participantes construirán un modelo para valuar propiedades en base al dataset de Properati.

Problemas de Clasificación, GridSearch, Pipelines y Procesamiento distribuido

04

- Intro a Clasificación y KNN
- Regresión Logística
- Support Vector Machines
- Naive Bayes Classifiers
- Evaluación de modelos
- Procesamiento Distribuido

Desafío del Módulo

Los participantes construyen un modelo para predecir la probabilidad de clicks dentro de aplicaciones móviles en base a un dataset de Jampp.

Text Mining,
Series de Tiempo

05

- Text Mining
- Series de Tiempo

Proyecto Integrador

El Proyecto Integrador (PI) debería representar un aporte original y significativo, aplicando técnicas de data science a un problema interesante.

Charla relámpago:

- Planteo del problema
- Selección de datasets

Árboles y Métodos de Ensamble

06

- Intro a CARTS
- Árboles de Decisión y Bagging
- Random Forests y Boosting
- XGBoost
- Evaluación de Modelos y Feature Importance

Proyecto Integrador

Informe de avance:

- Análisis Exploratorio
- Primeros intentos con el/los algoritmo(s) seleccionado(s)
- Resultados preliminares

Aprendizaje no supervisado: PCA, Clustering, Manifold

07

- Intro a Clustering
- K-means
- Intro a Clustering Jerárquico
- DBSCAN
- PCA
- Manifold Learning

Proyecto Integrador

Entrega Final

- Reporte técnico detallado con todos los análisis desarrollados (en formato notebook)
- Presentación de 10-15 minutos con los insights más relevantes del proyecto
  - Objetivos
  - Datasets
  - Métodos
  - Visualizaciones
  - Storytelling

DESAFÍOS Y PROYECTO INTEGRADOR

PROYECTO INTEGRADOR

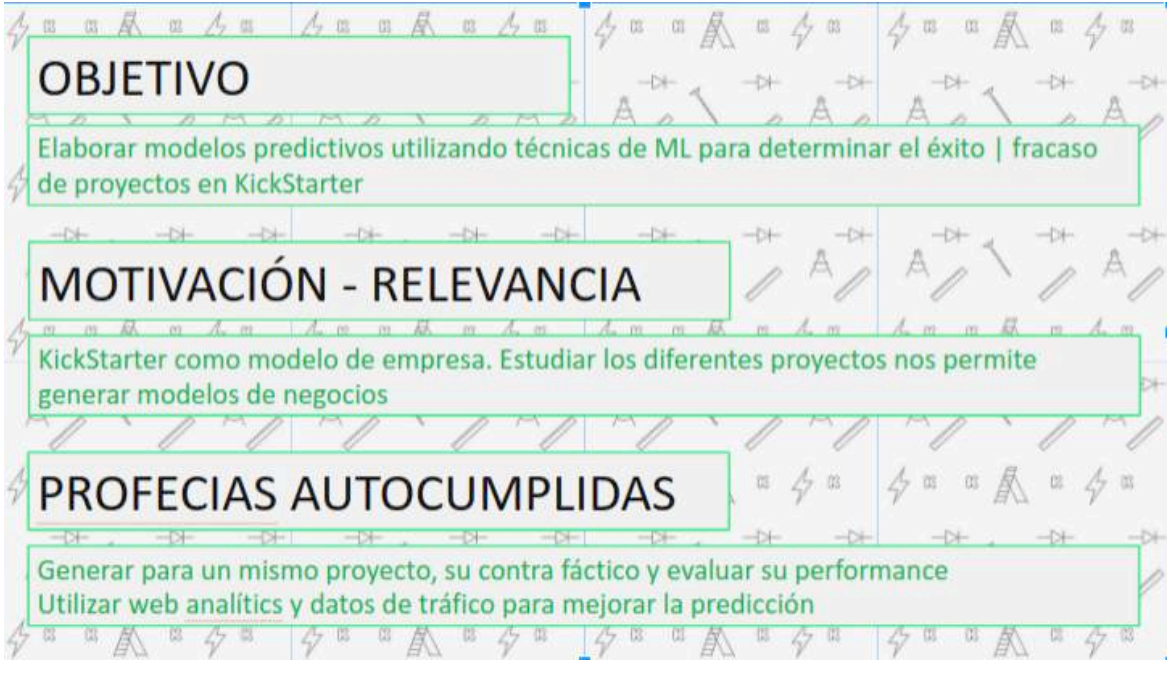
- **Desafíos y proyectos - objetivos generales:**
  - Resolver un problema práctico
  - Generar un reporte técnico (con código y análisis)
  - Generar un reporte para una audiencia no técnica
- **Desafíos Final: Proyecto Integrador: recorrer todo el Flujo de Trabajo de Data Science**
  - Planteo y fundamentación de un problema
  - Generación/adquisición de un dataset apropiado para el problema
  - Análisis, modelado y visualización de resultados
  - Presentación técnica y no técnica de hallazgos y conclusiones

ESCUCHA DE REDES SOCIALES PARA LA GESTIÓN PÚBLICA

Francisco PENSA

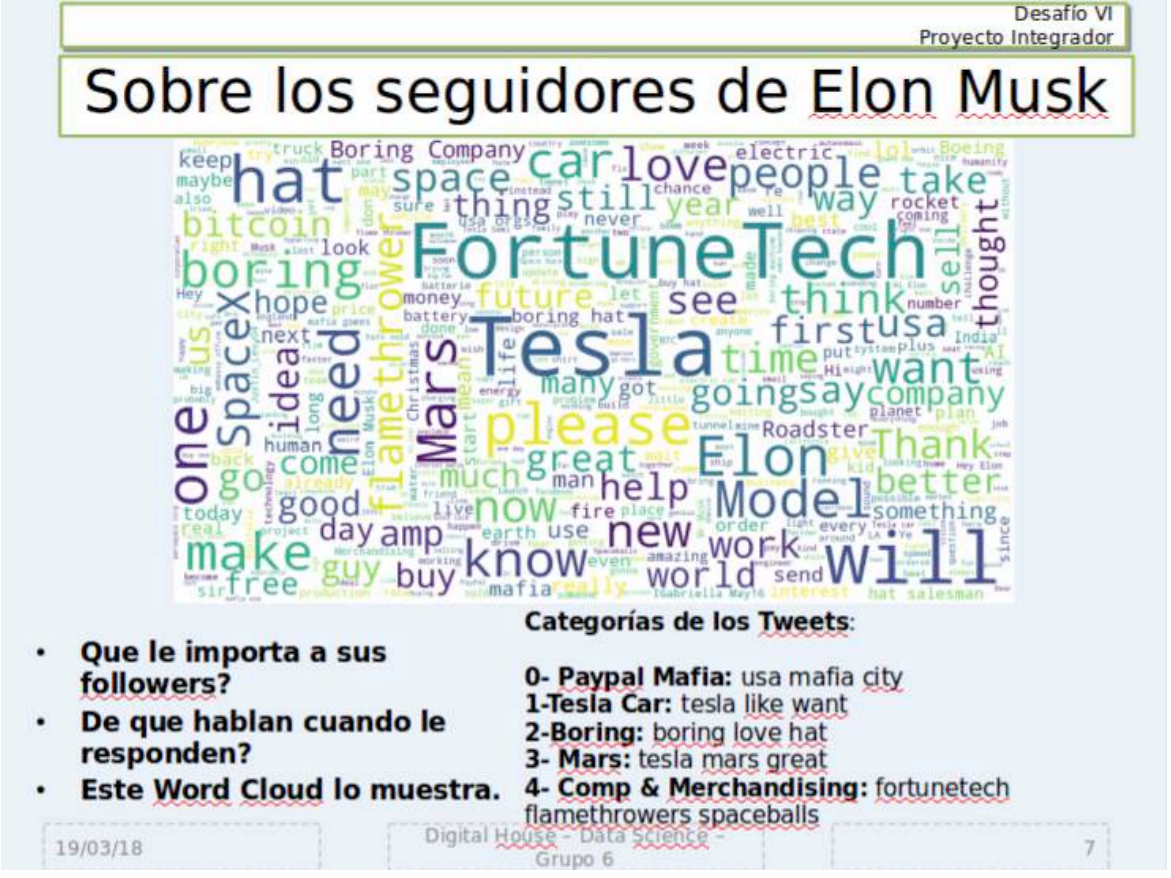
KICK-ASS MACHINE LEARNING: ¿QUÉ DETERMINA EL EXITO DE PROYECTOS EN LA PLATAFORMA KICKSTARTER?

José SANCHEZ, Jonathan COHEN



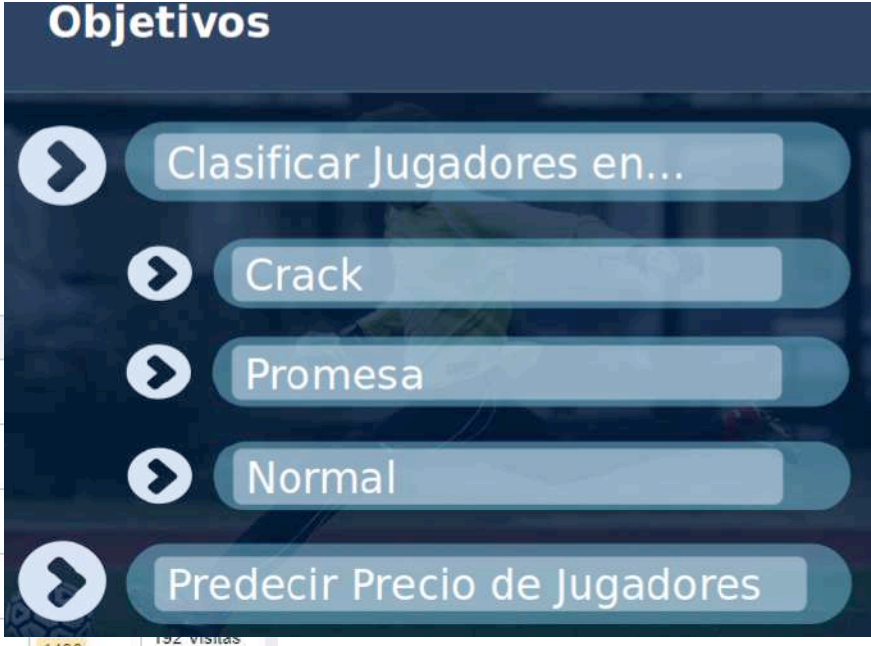
Sentiment Analysis y Topic Modeling en Twitter

Juan ARANGUREN
Jose Luis FINOCCHIARO



Promesas del Fútbol Mundial

Roberto DI LISIO
Guido BOZZANO
Benjamin BELLOT, Natalia MORAN



	NOMBRE	ED.	OVR	POT	EQUIPO & CONTRATO	VALUE	WAGE		
	Cristiano Ronaldo EI DC	32	94	94	Real Madrid Club d 2009 ~ 2021	$107M	$633K		
	L. Messi ED DC SD	30	93	93	Fútbol Club Barcel 2004 ~ 2021	$117.6M	$633K		
	Neymar EI	25	92	94	Paris Saint-Germai 2017 ~ 2022	$137.8M	$314K		
	L. Suárez DC	30	92	92	Fútbol Club Barcel 2014 ~ 2021	$108.6M	$571K		
	M. Neuer POR	31	92	92	FC Bayern Munich 2011 ~ 2021	$68.3M	$258K	1493	192 Visitas 1566 Comn
	De Gea POR	26	91	93	Manchester United 2011 ~ 2019	$83.4M	$330K	1463	370 Visitas 6732 Comn
	R. Lewandowski DC	28	91	91	FC Bayern Munich 2014 ~ 2021	$103M	$398K	2150	276 Visitas 3442 Comn
	K. De Bruyne MCO MC	26	90	92	Manchester City 2015 ~ 2021	$104.2M	$319K	2183	1640 Visitas 6128 Comn
	E. Hazard EI MCO	26	90	91	Chelsea 2012 ~ 2020	$101.4M	$330K	2109	406 Visitas 8232 Comn

DigitalHouse > Coding School

Conociendo a los participantes del programa usando Data Science (40 minutos)

DigitalHouse >
Coding School

Te proponemos

- Que todos los participantes del programa se conozcan mutuamente usando algunos pasos del Flujo de Trabajo de Data Science.
- Que formen grupos de 4 a 6 personas
- Que cada grupo defina **una** pregunta sobre algún aspecto que le interese conocer acerca de los compañeros (motivación, formación, etc.)
- Que a partir de la Encuesta Introductoria al curso puedan abordar las preguntas planteadas.

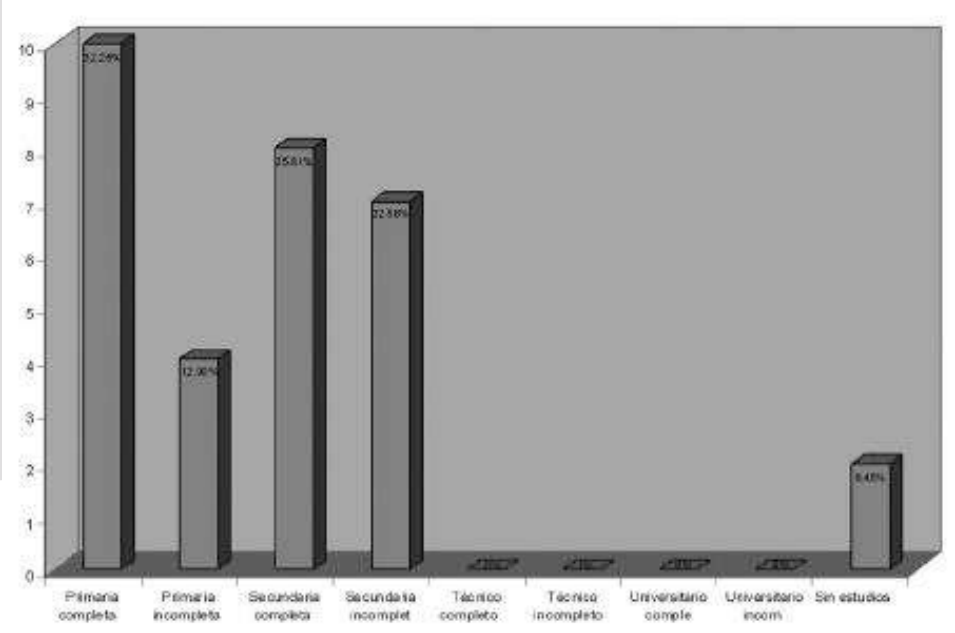
DigitalHouse >
Coding School

La idea es que...

- Cada grupo defina los siguientes roles:
  - 1 Project Manager (PM) - Data Business Person: responsable del cumplimiento de los tiempos, de facilitar la comunicación y hacer seguimiento del flujo de trabajo
  - 1 a 3 Researchers: encargados de adecuar la pregunta a los datos disponibles y de resumir la información para obtener la respuesta. Arman visualizaciones lo más claras y sintéticas posibles de la pregunta en cuestión
  - 1 a 2 Comunicadores-Creativos: encargados de resumir y presentar los hallazgos y conclusiones a los participantes

Por Ejemplo

- ¿Cuál es el perfil educativo del curso de Data Science-2017?
  - Primario incompleto
  - Primario completo
  - Secundario incompleto
  - Secundario completo
  - Universitario/Terciario incompleto
  - Universitario/Terciario completo
  - Posgrado o superior
  - Sin Estudios



DigitalHouse >
Coding School

Cronograma

Actividad	Tiempo	Responsable
Formación de grupos y distribución de roles	5 minutos	Equipo
Diseño de la pregunta	5 minutos	Equipo
Resumen y visualizaciones de la información	15 minutos	Analistas, Presentadores
Presentación de resultados	10 minutos	Presentadores

Al final del curso, ustedes serán capaces de:

- Extraer, consultar, limpiar y agregar datos para su análisis.
- Realizar análisis visuales y estadísticos de datos, usando Python y sus bibliotecas asociadas.
- Construir, implementar y evaluar problemas de Data Science usando los algoritmos apropiados de machine learning.
- Usar las herramientas de visualización adecuadas para comunicar sus conclusiones.

- Crear reportes claros y reproducibles para los stakeholders.

- Investigar, modelar y validar procesos de resolución de problemas aplicados a datasets provenientes de diversas industrias para proveer experiencias en distintos tipos de problemas y soluciones del mundo real.