



Módulo VIII: Deployando un modelo

Clase 31: Espacio práctico - Desarrollo de portfolio





**¿Ponemos a grabar
el taller?**

¿Con qué practicaremos?



- ¿Cómo desarrollo un portfolio?
- Dudas



Desarrollo de Portfolio

Importancia del portfolio en DS

Es la **mejor manera** de mostrar habilidades (Python, modelado de datos) a través de ideas interesantes de proyectos de ciencia de datos que se puedan implementar y compartir.

¿Cómo empezar?

- Elige temas que te gusten
- Busca datos relacionados que se puedan analizar
 - Trata de no utilizar los datasets más trillados y sencillos (Titanic, MNIST, Iris)
- Sumate a eventos: Hackathons, Competencias, etc.

Datasets

kaggle

Google
Dataset Search Beta



... Y en otros bancos de datos
abiertos.

Datasets

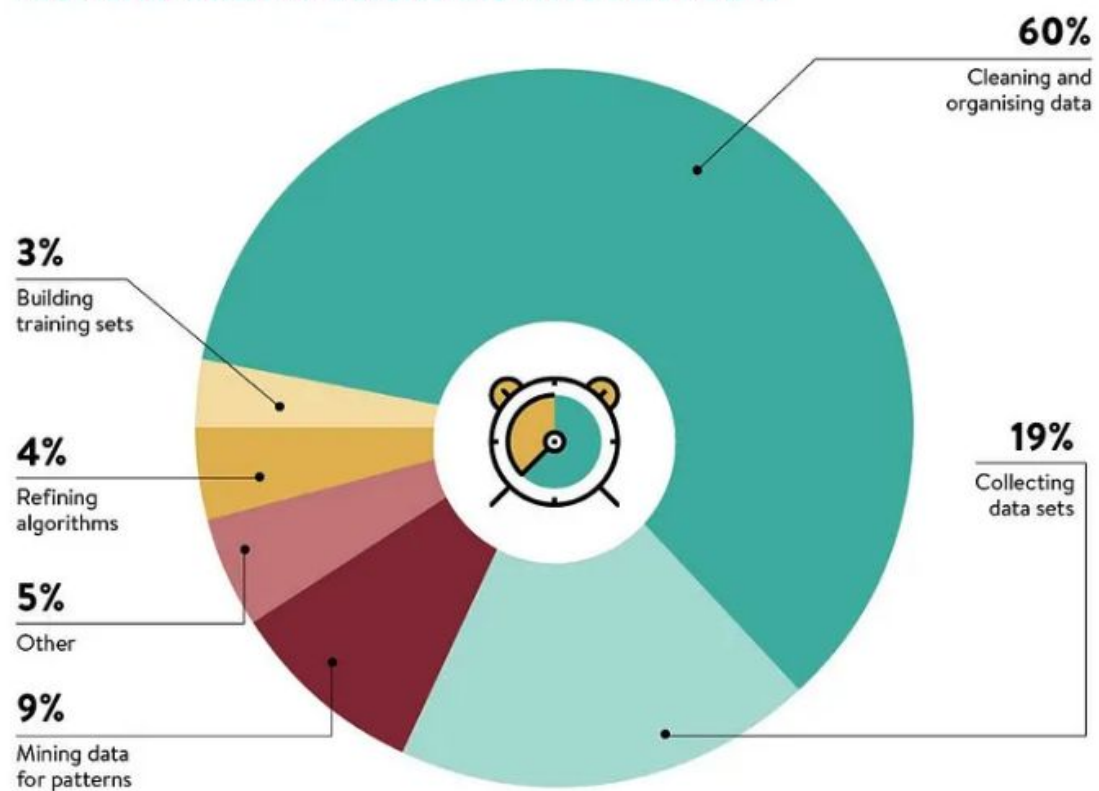
O podemos crear nuestro propio dataset



 Developer Platform

¿Qué trabajo mostrar?

WHAT DATA SCIENTISTS SPEND THE MOST TIME DOING



Source: CrowdFlower 2016

Imagen de CrowdFlower 2016.

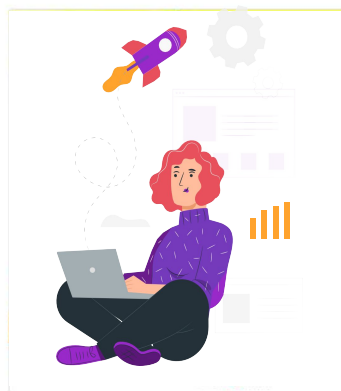
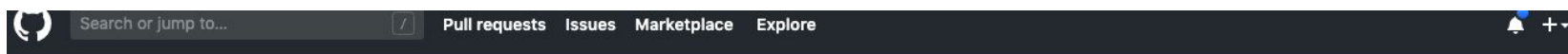
¿Qué trabajo mostrar?

- Recolección de datos
 - Web Scraping
- Limpieza de datos
- Exploración y Análisis de Datos
- Modelado y Predicción de nuevos ejemplos
 - Regresión y Clasificación
 - Sistemas de Recomendación
 - Análisis de Sentimiento
 - Clustering y Segmentación de Clientes
 - Otros ...

¿Dónde mostrar mi trabajo?



GitLab



Eugenia Inzaugarat
ugis22



Edit profile

• Data Scientist •
<https://www.linkedin.com/in/meinzaugarat>
Argentina

Overview

Repositories 23

Projects 0

Stars 2

Followers 13

Following 2

Pinned

Customize your pins

[clustering_analysis](#)

Performs an exploratory analysis on a dataset containing information about shop customers. Check that the assumptions K-means makes are fulfilled. Apply K-means clustering algorithm in order to seg...

Jupyter Notebook ★ 2 🍴 1

[creatingDCGAN](#)

Show how DCGAN works. Contains some code about the generator, discriminator and training step.

Python

[analysing_twitter](#)

Stream Tweets and store them in a relational DB. Perform sentiment analysis and network interaction.

Jupyter Notebook ★ 7 🍴 1

[essential_tools_ner](#)

Retrieves information from a particular website. Uses different tools and pre-trained models for Named Entity Recognition.

Jupyter Notebook

[movie_predictions](#)

Analyze a dataset with information about movies. Creates a linear and a bayesian model to predict movie popularity.

HTML ★ 2 🍴 2

[predicting_survival_hepatitis](#)

Analyze a dataset with information about liver disease patients. Creates a model to predict their survival.

Jupyter Notebook

¿Dónde mostrar mi trabajo?

≡ kaggle

+ Create

🏠 Home

🏆 Competitions

📊 Datasets

🤖 Models

<> Code

💬 Discussions

🎓 Learn

🔍 Search

Code

Explore and run machine learning code with Kaggle Notebooks.
Find help in the [Documentation](#).

+ New Notebook

Your work

🔍 Search public notebooks

All notebooks

Recently Viewed

Python

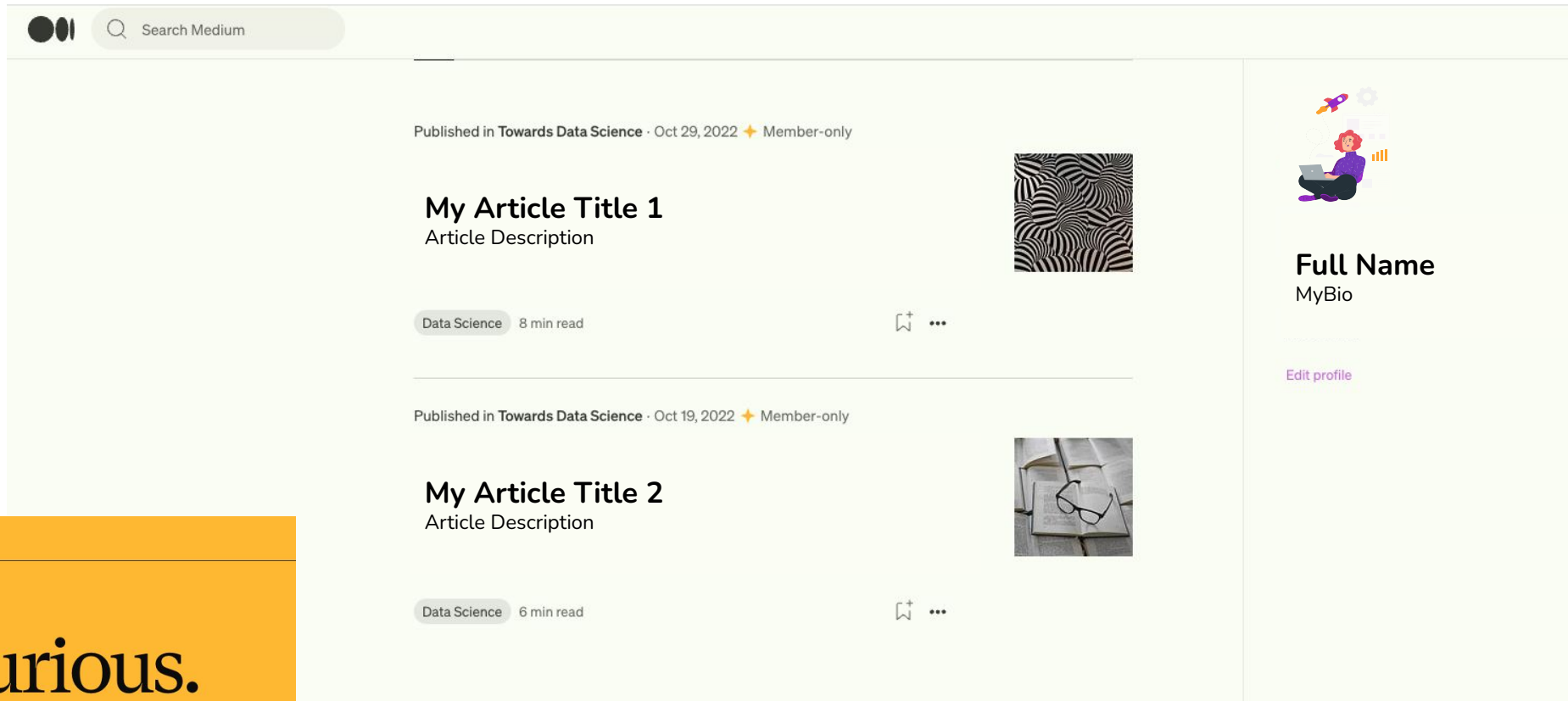
R

Beginner

NLP

Random Forest

¿Dónde mostrar mi trabajo?



Medium

Stay curious.

Discover stories, thinking, and expertise
from writers on any topic.

Habilidades a demostrar

- Habilidades en programación
- Entendimiento de los problemas
- Búsqueda de respuestas usando datos
- Conocimiento TEÓRICO de cómo funcionan los modelos (A través de selección de hiperparametros, transformación de datos, manejo de datos faltantes)
- Storytelling y Comunicación (Al describir trabajo, escribir artículos y contar una historia con datos)



Descanso

Nos vemos en 10 minutos

Trabajamos en salas

Avanzamos con la entrega final

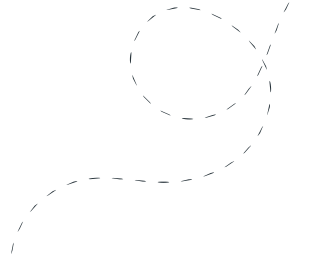
En los grupos establecidos, avanzamos reuniendo los desafíos solicitados y armando una presentación.

Preparando la entrega final

En la **última clase** deberán presentar un resumen de lo trabajado durante el curso. Para ello podrán preparar una **presentación** en PowerPoint, Slides o Canva.

Tendrán 15 minutos para contar a sus compañeras la síntesis de lo aprendido.

Además, para recibir el certificado, deberán adjuntar en el aula virtual un link a su repositorio. El mismo debe incluir los ítems detallados en la siguiente hoja.



Preparando la entrega final

Ítems a incluir en el repositorio:

- Readme (resumen, no más de un párrafo por ítem. En el mismo pueden contar de dónde obtuvieron dataset, resumen proyecto, qué objetivo se propusieron y qué descubrieron en análisis exploratorio, detalle de los modelos ajustados)
- Dataset
- Notebook con análisis exploratorio (correspondiente a pre-entrega 2)
- Notebooks donde apliquen modelos de aprendizaje supervisado y no supervisado y donde se observen resultados obtenidos (correspondientes a pre-entrega 3 y 4)
- Será valorado que cuente una historia con esos datos: que puedan demostrar cualitativamente qué preguntas se hicieron, conclusiones que sacaron con los datos, decisiones tomadas

¿Dudas? ¿Consultas?

FUNDACIÓN
YPF

¡Muchas gracias!

