

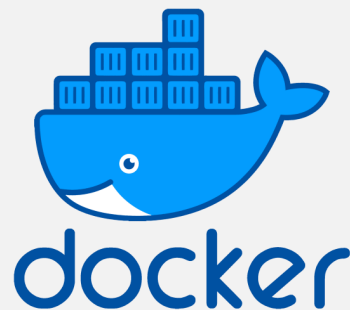


DESPLIEGE DE ML CON R: CREACIÓN DE PAQUETES Y API

Jose Vicente Yago Martinez



VicenteYago/deployingMLinR



PARTE 1: Construir un paquete en R

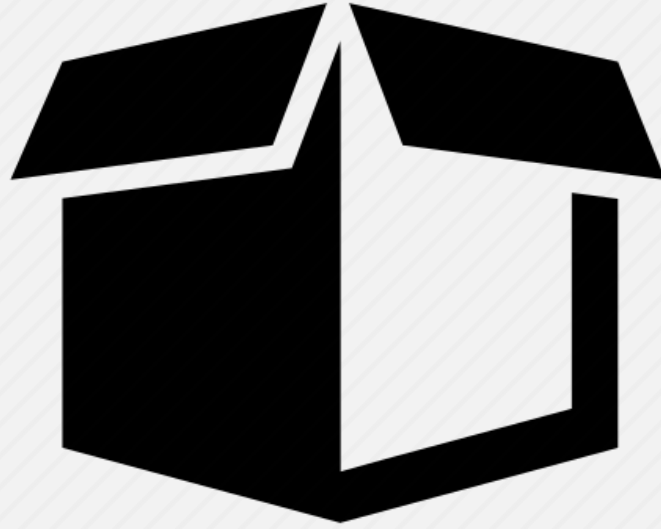
- ¿Qué es un paquete en R?
- Otras definiciones
- Ejemplos reales
- Manos a la obra!

PARTE 1: Construir un paquete en R

- ¿Qué es un paquete en R?
- Otras definiciones
- Ejemplos reales
- Manos a la obra!

PARTE 2: Despliegue de modelos ML

- ¿Por qué es necesario?
- Herramientas útiles
 - Bash
 - Docker
- ¿Tengo que hacer un paquete para desplegar mi modelo de ML?
- Alternativas existentes
 - Plumber
 - OpenCPU
- Manos a la obra!



PARTE I: CONSTRUIR UN PAQUETE EN R

¿Qué es un paquete en R?

¿Qué es un paquete en R?

- Unidad fundamental de código compartible

¿Qué es un paquete en R?

- Unidad fundamental de código compartible
- Reune código, datos, documentación y pruebas.

¿Qué es un paquete en R?

- Unidad fundamental de código compartible
- Reune código, datos, documentación y pruebas.
- Enorme variedad de paquetes

¿Qué es un paquete en R?

- Unidad fundamental de código compartible
- Reune código, datos, documentación y pruebas.
- Enorme variedad de paquetes
- "Si necesitas hacer algo, posiblemente haya un paquete R que ya lo implemente."

¿Qué es un paquete en R?

- Unidad fundamental de código compartible
- Reune código, datos, documentación y pruebas.
- Enorme variedad de paquetes
- "Si necesitas hacer algo, posiblemente haya un paquete R que ya lo implemente."
- **Comprehensive R Archive Network (CRAN)**

Otras definiciones

- **Package repository:** lugar donde se almacenan paquetes -> CRAN, GitHub

Otras definiciones

- **Package repository:** lugar donde se almacenan paquetes -> CRAN, GitHub
- **Source package:** directorio que contiene ficheros de paquete para su desarrollo

Otras definiciones

- **Package repository:** lugar donde se almacenan paquetes -> CRAN, GitHub
- **Source package:** directorio que contiene ficheros de paquete para su desarrollo
- **Binary package:** archivo binario específico de un SO + arquitectura, que contiene el **source package** compilado.

Otras definiciones

- **Package repository:** lugar donde se almacenan paquetes -> CRAN, GitHub
- **Source package:** directorio que contiene ficheros de paquete para su desarrollo
- **Binary package:** archivo binario específico de un SO + arquitectura, que contiene el **source package** compilado.
- **Base package:** paquete esencial de R para su funcionamiento. Se cargan automáticamente en cualquier environment.

Ejemplos reales

- <https://github.com/robjhyndman/forecast>
- <https://github.com/tidyverse/dplyr>
- <https://github.com/ranghetti/sen2r>

Manos a la obra!

Paquete dummyML :

- Regresión lineal simple (base R)
- Random forests (tidymodels)

Manos a la obra!

Paquete dummyML :

- Regresión lineal simple (base R)
- Random forests (tidymodels)
- Documentación básica
- Gestión de dependencias
- Test unitarios



PARTE 2: DESPLIEGUE

¿Desplegar un modelo?

¿Desplegar un modelo?

- El trabajo de un Data Scientist / ML Engineer no termina cuando su modelo alcanza un ACC determinado.

¿Desplegar un modelo?

- El trabajo de un Data Scientist / ML Engineer no termina cuando su modelo alcanza un ACC determinado.
- Un modelo ML solo es útil para los clientes cuando se puede ejecutar de forma rutinaria

¿Desplegar un modelo?

- El trabajo de un Data Scientist / ML Engineer no termina cuando su modelo alcanza un ACC determinado.
- Un modelo ML solo es útil para los clientes cuando se puede ejecutar de forma rutinaria
 - No se ejecutará en tu pc

¿Desplegar un modelo?

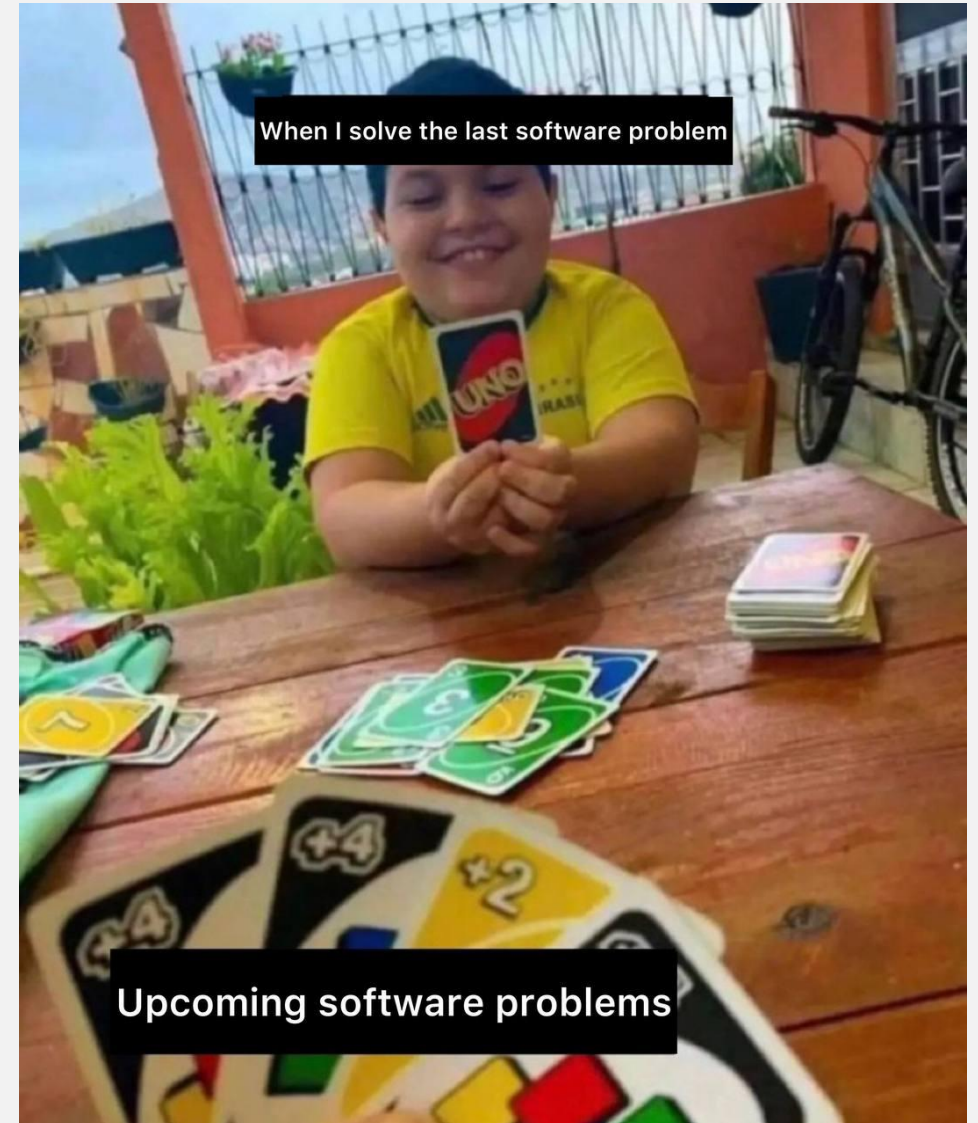
- El trabajo de un Data Scientist / ML Engineer no termina cuando su modelo alcanza un ACC determinado.
- Un modelo ML solo es útil para los clientes cuando se puede ejecutar de forma rutinaria
 - No se ejecutará en tu pc
 - No se alimentará de CSVs

¿Desplegar un modelo?

- El trabajo de un Data Scientist / ML Engineer no termina cuando su modelo alcanza un ACC determinado.
- Un modelo ML solo es útil para los clientes cuando se puede ejecutar de forma rutinaria
 - No se ejecutará en tu pc
 - No se alimentará de CSVs
 - El output no será a través de un notebook/terminal

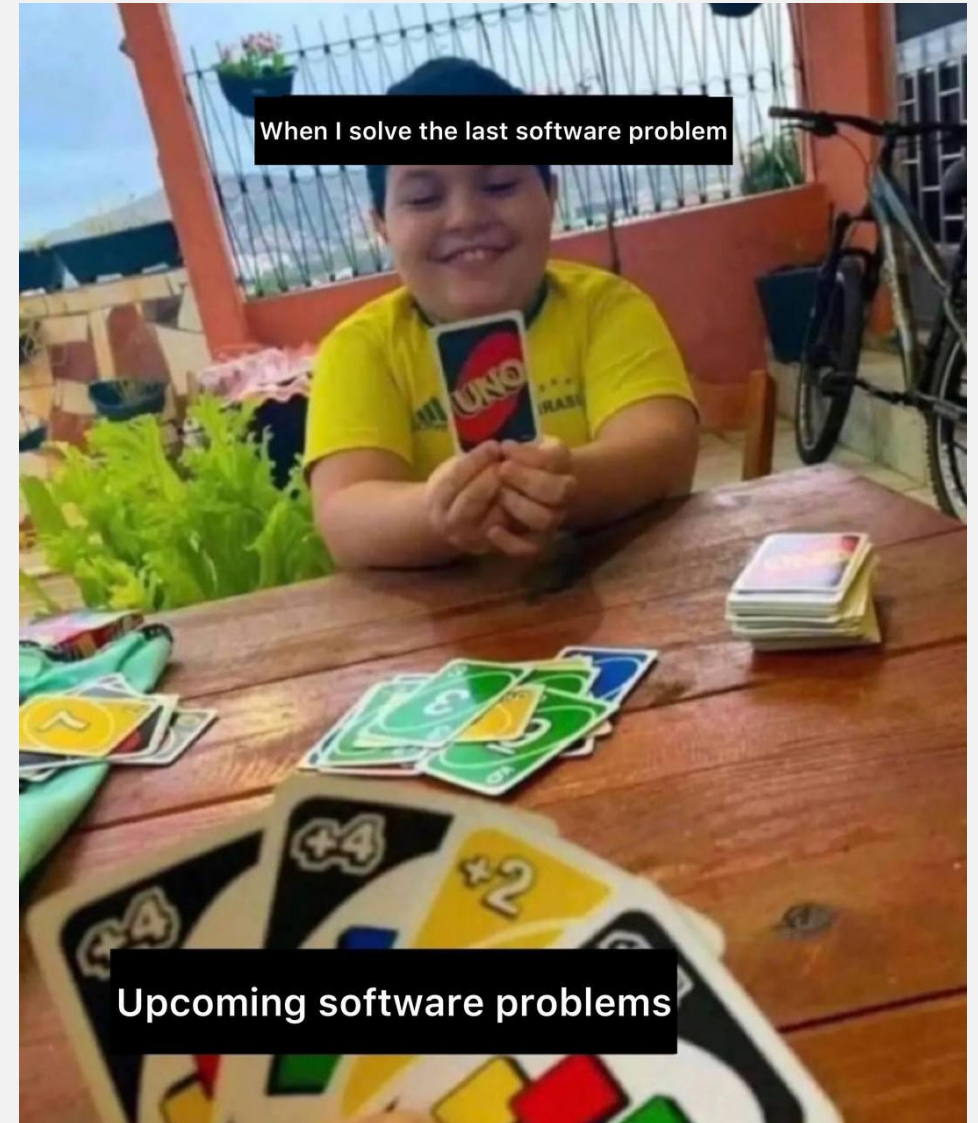
¿Desplegar un modelo?

- Puede ser complicado si no se tiene nociones de desarrollo de software.



¿Desplegar un modelo?

- Puede ser complicado si no se tiene nociones de desarrollo de software.
- Y si se tienen, probablemente también



¿Tengo que hacer un paquete para desplegar mis modelos de ML ?

¿Tengo que hacer un paquete para desplegar mis modelos de ML ?

Agrupas los modelos en un paquete facilita:

- Mantenimiento
- Desarrollo
- Seguimiento

No es necesario, pero si muy útil

Alternativas existentes

Alternativas existentes

PLUMBER



Alternativas existentes

PLUMBER

OPENCPU



Alternativas existentes

	FACIL	DOCS	ESCALABLE	SEGURIDAD	CUSTOM	COMUNIDAD
PLUMBER						
OPENCPU						

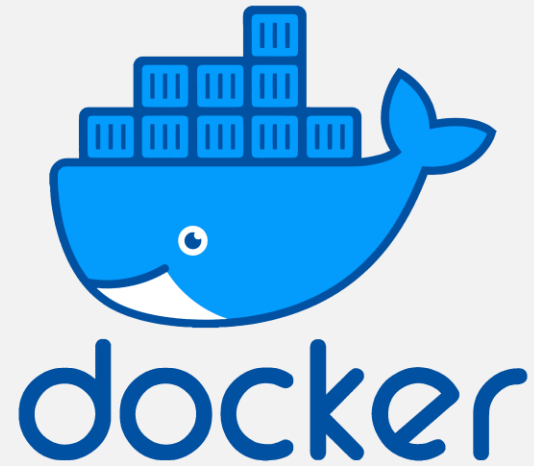
Alternativas existentes

	FACIL	DOCS	ESCALABLE	SEGURIDAD	CUSTOM	COMUNIDAD
PLUMBER						
OPENCPU						

- OpenCPU no es solo un paquete, es una interfaz completa a R, integrada en un servidor Apache, con todas las ventajas de escalabilidad y seguridad que conlleva.

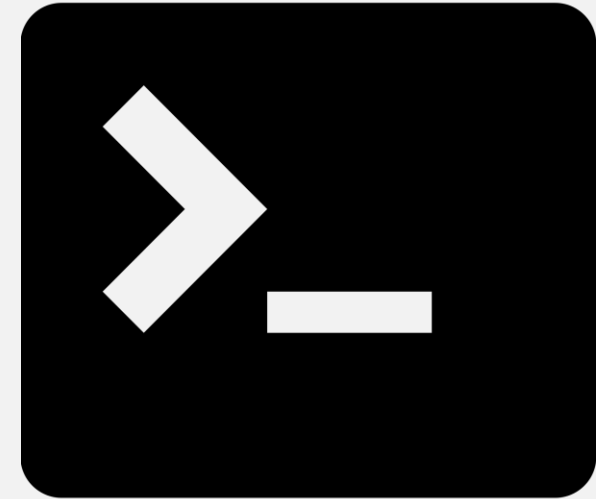
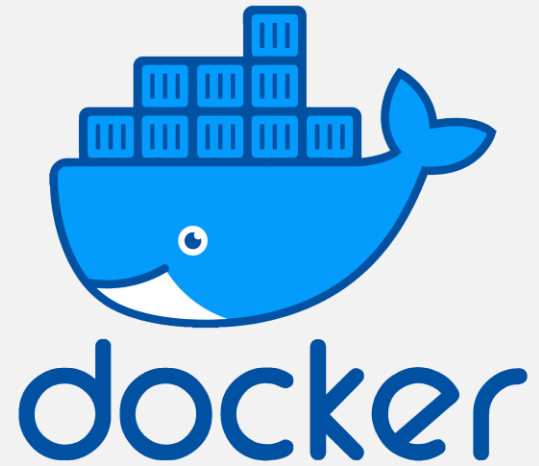
Herramientas útiles

Herramientas útiles



<https://www.rocker-project.org/>

Herramientas útiles



<https://www.rocker-project.org/>

Manos a la obra!

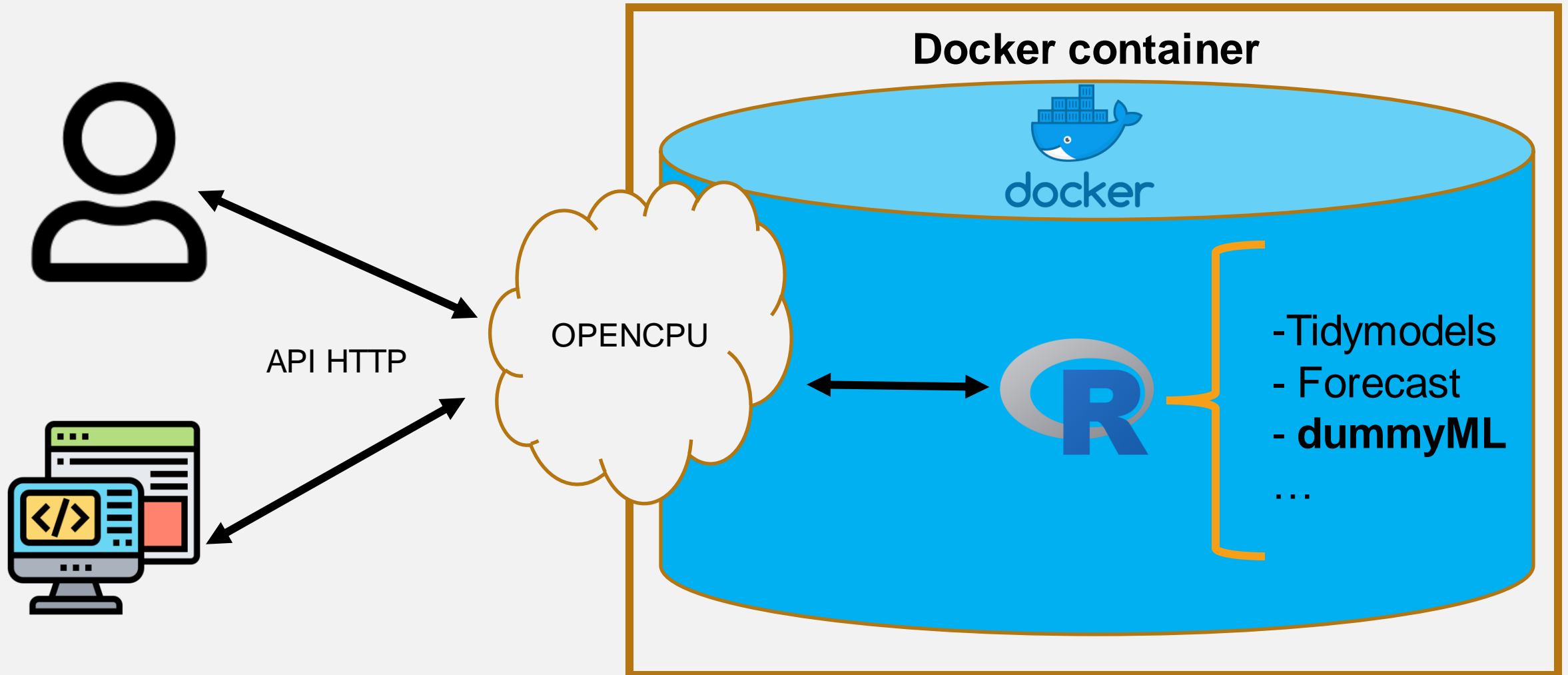
Manos a la obra!

- Familiarizarnos con `opencpu`, `bash` y `docker`.
- Despliegue en OpenCPU del paquete `dummyML`

Arquitectura

Arquitectura

HOST (server aka mi pc)



¡GRACIAS!