

Presentación de Orange



Ezequiel Medina

FaMAF(UNC) - Machinalis



Qué es Orange?

Orange es una herramienta open source de **visualización de datos**, **aprendizaje automático** y **minería de datos** programada en Python.

Podemos usar Orange como **biblioteca de Python** o podemos utilizar su interfaz gráfica que nos permite **programar de forma visual**.

La primera versión fue creada en **1997** en el laboratorio de bioinformática de la facultad de ciencias de la información en **Eslovenia**.



Por qué Orange?

Es muy útil para hacer un primer **análisis exploratorio** sobre los datos con los que vamos a trabajar.

Es muy potente, permite correr **algoritmos de ML** e incluso **redes neuronales**.

Es ideal para personas que trabajan con datos y no son programadores.



Qué podemos hacer con Orange?

- **Procesamiento de datos**
 - Preprocesamiento (filtrar, modificar, sacar outliers)
 - Construcción de Features
 - Estadísticos
- **Visualización de datos**
 - Box plot, Scatter plot
 - Distribuciones
 - Heatmaps
 - Clusters
- **Algoritmos de aprendizaje**
 - Clasificadores
 - Regresores
 - Confusion Matrix
 - ROC Analysis
 - Feature Scoring

Y más..



Problema: Titanic Survival Predictions

Variables del Dataset

Survival:	0 = No, 1 = Yes
Pclass:	1 = 1st, 2 = 2nd, 3 = 3rd
Sex:	male, female
Age:	Años
Sibsp:	# de hermanos / cónyuges a bordo
Parch:	# de padres / hijos a bordo
Ticket:	Número de Ticket
Fare:	Precio del pasaje
Cabin:	Número de cabina
Embarked:	Puerto donde embarcó (C = Cherbourg, Q = Queenstown, S = Southampton)

Preguntas?

