

Interpretación de
Modelos Predictivos
mediante
Valores Shapley

Nuestro objetivo último sería a partir de datos sobre una realidad compleja (y quizás de un poco de experimentación) lograr un modelo matemático causa-efecto, modelo de causalidad, que nos calcule el valor de las variables de salida ante intervenciones en las variables de entrada, y por lo tanto permita optimizar el sistema, “rentabilizar el negocio”.

Sería muy deseable que este modelo de causalidad sirviera para valores de entrada que están fuera de los rangos presentes en los datos de entrada.

Todos los modelos predictivos que tenemos, desde una regresión lineal hasta Gradient Boosting y Deep Learning solamente encuentran correlaciones entre los valores de entrada y los de salida.

Correlación no implica causalidad.

NO ESTAMOS RESOLVIENDO EL
PROBLEMA DE LA CAUSALIDAD.

Apenas vamos a hechar un poco más de luz sobre las correlaciones, encontrando explicaciones *locales* sencillas a los modelos predictivos, y aproximaciones globales.

Un gran problema del data mining es :

- Modelos como la regresión lineal

$$y = a_1x_1 + a_2x_2 + \dots + a_nx_n + b$$

tienen explicaciones aditivas simples, en donde se observa claramente el aporte de cada variable al resultado final; pero tienen un poder predictivo bajo.

Debe aclararse que el poder predictivo es bajo porque en los datos existen relaciones NO LINEALES entre los atributos x con la clase y

por otro lado

- Modelos con gran poder predictivo como Gradient Boosting tienen el estado del arte en poder predictivo al capturar las relaciones no lineales entre los atributos PERO no es humanamente entendible combinación de 300 arboles de altura 6 ...

No es facil saber como contribuye cada una de las variables al resultado final.

Resultados de lgb.importance()

Gain	Cover	Frequency
mcaja_ahorro_Paquete	Master_Fvencimiento__tend	mplan_sueldo__tend
tmovimientos_ultimos90dias	mv_msaldodolares__min	canarito46
tmovimientos_ultimos90dias__min	cliente_antiguedad__min	mcuentas_saldo__tend
mv_cuenta_estado2	tcajas_extracciones__min	mcuenta_corriente_Paquete
Master_cuenta_estado__tend	mvr_mconsumospesos__tend	mprestamos_personales__min
mcuentas_saldo__tend	Master_madelantodolares__max	canarito50

Dado un modelo predictivo M , y un nuevo registro r , sea $p = M(r)$ la probabilidad que M asigna a r de ser positivo.

¿ Existe una explicación de que al registro r el modelo M le asignó la probabilidad p ?

¿ Qué variables jugaron a favor y cuales en contra de la probabilidad p para el registro r ?

¿Cuál fue la contribución de cada variable?

Teoría de Juegos

Juegos Cooperativos



Lloyd Shapley, 1923-2016
Economic Sciences Nobel Prize, 2012

How to bring different players together in the best possible way is a key economic problem.

From the 1960s onward, Lloyd Shapley used what is known as Cooperative Game Theory to study different matching methods. Within the framework of this theory, it is especially important that a stable match is found.

A stable match entails that there are no two agents who would prefer one another over their current counterparts.

In collaboration with other researchers, Lloyd Shapley has succeeded in identifying methods that achieve this stability.

Motivación

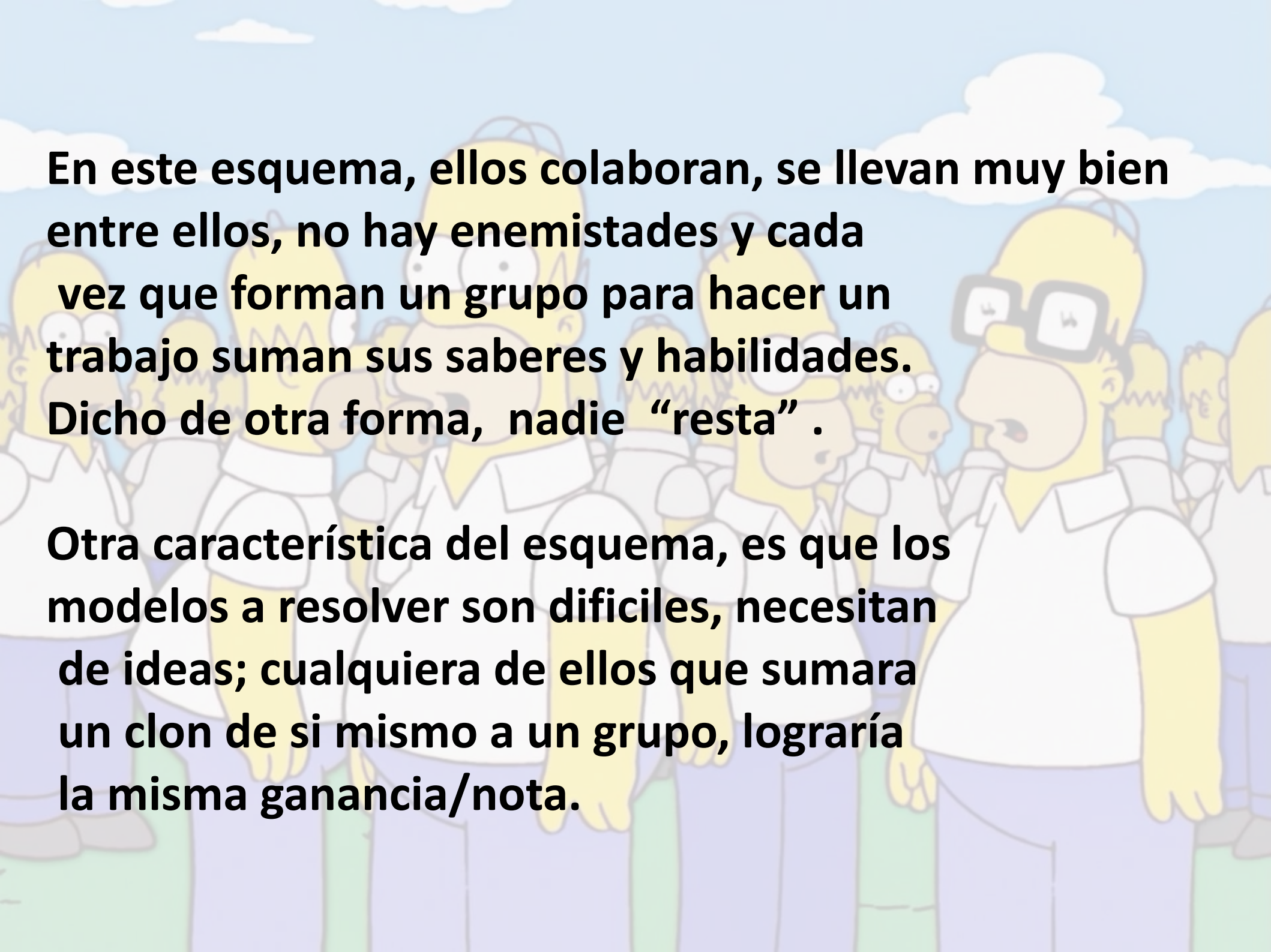
En un país muuuuy lejano, Alexandre, Francesco y Luigi cursan una maestría en Ciencia de Datos. Para una de las materias debieron hacer varios modelos donde los median por la ganancia obtenida.

Algunos modelos los debieron de hacer cada uno por su lado individualmente, otros colaborando en grupos de a dos y hubo un caso, cuando trabajaron los tres juntos que obtuvieron un puntaje de 95/100 .

Motivación

Cada uno tiene sus fortalezas y debilidades para la confección de modelos predictivos, producto de su formación de grado, experiencia profesional previa, afinidad con distintas etapas del data mining y esfuerzo dedicado.

Algunas habilidades son compartidas entre los tres, otras compartidas de a dos en distintos grados, otros saberes son exclusivos de cada uno y agregan mucho al grupo, y por qué no decirlo, también hay carencias comunes a los tres, (son las que evitaron el 100/100)



En este esquema, ellos colaboran, se llevan muy bien entre ellos, no hay enemistades y cada vez que forman un grupo para hacer un trabajo suman sus saberes y habilidades. Dicho de otra forma, nadie “resta” .

Otra característica del esquema, es que los modelos a resolver son difíciles, necesitan de ideas; cualquiera de ellos que sumara un clon de si mismo a un grupo, lograría la misma ganancia/nota.

Grupo	Nota
{ A }	48
{ F }	31
{ L }	52
{ A, F }	71
{ A, L }	65
{ F, L }	83
{ A, F, L }	95

¿ Dadas las notas obtenidas por cada uno de los grupos, como se podría medir el aporte de Alexandre, Francesco y Luigi al grupo $\{ A, F, L \}$?

Propiedades Deseables

- El jugador nulo contribuye 0
- Eficiencia, la suma de las todas contribuciones individuales mas una constante es igual a la contribución total (esto no se cumple parcialmente)
- Simetria, si dos alumnos son “equivalentes” entonces su contribucion debe ser la misma
- Linearidad, $\phi(v+w)=\phi(v)+\phi(w)$

Propiedades Deseables

Lloyd Shapley demostró en 1951 para su tesis de doctorado que hay una sola forma de asignar contribuciones a los integrantes de un equipo tal que se cumplan las cuatro propiedades deseables anteriores. Esos valores únicos para cada integrante se llaman *valores Shapley*

Valores de Shapley

$$\varphi_i(v) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|! (n - |S| - 1)!}{n!} (v(S \cup \{i\}) - v(S))$$

$$\varphi_i(v) = \frac{1}{\text{number of players}} \sum_{\text{coalitions excluding } i} \frac{\text{marginal contribution of } i \text{ to coalition}}{\text{number of coalitions excluding } i \text{ of this size}}$$

Metodología cálculo Valores Shapley

Permutaciones Ordenadas
contribución marginal

Grupo	Nota
{ A }	48
{ F }	31
{ L }	52
{ A, F }	71
{ A, L }	65
{ F, L }	83
{ A, F, L }	95

Permutac	Contribución		
	A	F	L
(A, F, L)	48		
(A, L, F)			
(F, A, L)			
(F, L, A)			
(L, A, F)			
(L, F, A)			
Promedio			

Grupo	Nota
{ A }	48
{ F }	31
{ L }	52
{ A, F }	71
{ A, L }	65
{ F, L }	83
{ A, F, L }	95

Permutac	Contribución		
	A	F	L
(A, F, L)	48	23	
(A, L, F)			
(F, A, L)			
(F, L, A)			
(L, A, F)			
(L, F, A)			
Promedio			

$$23 = 71 - 48$$

Grupo	Nota
{ A }	48
{ F }	31
{ L }	52
{ A, F }	71
{ A, L }	65
{ F, L }	83
{ A, F, L }	95

Permutac	Contribución		
	A	F	L
(A, F, L)	48	23	24
(A, L, F)			
(F, A, L)			
(F, L, A)			
(L, A, F)			
(L, F, A)			
Promedio			

$$24 = 95 - (23 + 48)$$

Grupo	Nota
{ A }	48
{ F }	31
{ L }	52
{ A, F }	71
{ A, L }	65
{ F, L }	83
{ A, F, L }	95

Permutac	Contribución		
	A	F	L
(A, F, L)	48	23	24
(A, L, F)	48		
(F, A, L)			
(F, L, A)			
(L, A, F)			
(L, F, A)			
Promedio			

Grupo	Nota
{ A }	48
{ F }	31
{ L }	52
{ A, F }	71
{ A, L }	65
{ F, L }	83
{ A, F, L }	95

Permutac	Contribución		
	A	F	L
(A, F, L)	48	23	24
(A, L, F)	48		17
(F, A, L)			
(F, L, A)			
(L, A, F)			
(L, F, A)			
Promedio			

$$17 = 65 - 48$$

Grupo	Nota
{ A }	48
{ F }	31
{ L }	52
{ A, F }	71
{ A, L }	65
{ F, L }	83
{ A, F, L }	95

Permutac	Contribución		
	A	F	L
(A, F, L)	48	23	24
(A, L, F)	48	30	17
(F, A, L)			
(F, L, A)			
(L, A, F)			
(L, F, A)			
Promedio			

$$30 = 95 - (17 + 48)$$

Grupo	Nota
{ A }	48
{ F }	31
{ L }	52
{ A, F }	71
{ A, L }	65
{ F, L }	83
{ A, F, L }	95

Permutac	Contribución		
	A	F	L
(A, F, L)	48	23	24
(A, L, F)	48	30	17
(F, A, L)	40	31	24
(F, L, A)	30	31	34
(L, A, F)	30	13	52
(L, F, A)	12	31	52
Promedio	34.6	26.5	33.8

Valores Shapley

Supongamos ahora que 16 alumnos cursan la materia,
¿Cuál es el valor Shapley de Francesco ?

Es el promedio de aumento de nota que genera que Francesco se sume a un equipo ya formado, es decir contribución marginal (equipos que se crean en orden), desde el equipo vacío hasta todos los equipos ordenados de 15 alumnos.

Aplicación de Valores Shapley a la importancia de variables

En 2017, Scott Lundberg aplicó los valores Shapley a la asignación de importancia de variables de un modelo predictivo.

Genera valores Shapley locales para cada predicción, y además un promedio general para cada atributo promediando sobre todo el dataset, que lo llama contribucion de esa variable.

La metodología es aplicable a cualquier modelo, pero para métodos basados en árboles corre rápido ya que no necesita calcular permutaciones, ni siquiera un muestreo montecarlo.