

Mayo 2020



Sataturdays.AI

This model fits me
95% of the time

Sataturdays.AI

Grupo Absolute Beginners

Participantes: Lizeth Ordóñez G. | Dirceo Bravo F. | Líder Técnico: Eric Garza (¡GRACIAS!) 🤖 😊 🧐 🙌



Contenido

Problemática

**Solución
Propuesta**

**¿Dónde
estamos?**

**Siguientes
pasos**

**¿Qué
aprendimos?**

**Saturday IA Absolute
Beginner Road Trip!**

Al principio no entendíamos todo lo que se presentaba



Las computadoras son increíblemente...
rápidas, precisas y estúpidas.
Por otro lado un usuario bien entrenado es
increíblemente...
lento, descuidado y altamente
inteligente.
Cuando la inteligencia humana se acopla
con la velocidad y precisión de una
computadora el resultado sobrepasa todo lo
que en el pasado era posible.

< Notes



Done

13 May 2020, 23:13

I.A. Saturday AI

Compañero musical
Consultor Terapeuta
Automatizar (predicción)





PREDECIR | AUTOMATIZAR | SUGERIR

...pero el Eric nos iba ayudando a digerir los temas
y con eso íbamos pensando en proyectos.



Principales temas cubiertos durante el curso



% del curso	Tema	Principal logro de aprendizaje
 25%	<ul style="list-style-type: none">• Programación básica en Python<ul style="list-style-type: none">◦ sintaxis básica◦ condicionales◦ bucles◦ etc.	<ul style="list-style-type: none">• Ejercicios de programación utilizando Google Colab
 50%	<ul style="list-style-type: none">• Álgebra lineal<ul style="list-style-type: none">◦ Vectors and Spaces◦ Matrix Transformations modules	<ul style="list-style-type: none">• Ejercicios utilizando Google Colab
 75%	<ul style="list-style-type: none">• Introducción a Machine Learning<ul style="list-style-type: none">◦ Regresión lineal◦ Generalización◦ Entrenamiento y prueba◦ Regularización◦ Regresión logística	<ul style="list-style-type: none">• Regresión - Regresión lineal• Clasificación - Regresión logística como un modelo de solución muy utilizado
 99%	<ul style="list-style-type: none">• Machine Learning<ul style="list-style-type: none">◦ Clasificación◦ Redes neuronales: entrenamiento, multiclase, backpropagation◦ Embeddings	<ul style="list-style-type: none">• Redes neuronales como modelo de solución muy utilizado• Ejemplo de cómo organizar películas por similitud, como Netflix

Posibles proyectos



1. **Clasificación automática de tickets** recibidos por HelpDesk

- Restricción: Se descartó porque los datos de la compañía son confidenciales y no deben extraerse ni cargarse a Google Colab.
- Avance: Tutor recomendó un ejemplo de Kaggle
<https://www.kaggle.com/aniketg11/support-tickets-classification>
- Siguientes pasos: Se intentará implementar este proyecto en el corto plazo dentro de la red de la compañía.

2. Predicción del éxito de una campaña de **Marketing en Facebook** (Seleccionado)



Proyecto: Análisis de un caso de uso de Marketing Digital



Solución Propuesta: Practicar un modelo de regresión para analizar los resultados de anuncios que facebook entrega. Realizados por una comunidad a través de su página

¿Cómo? Encontrando que atributos en la configuración de un anuncio pueden predicen un menor **Costo por Resultado** en el objetivo de los eventos o meetups!

“Datos y problemática para entender”

- Enfoque general, describir el problema de datos
- Progreso del problema

Contenido

Problemática

**Solución
Propuesta**

**¿Dónde
estamos?**

**Siguientes
pasos**

**¿Qué
aprendimos?**

Proyecto: Solución Propuesta



Objetivo: Predecir el valor de un campo o Feature que pueda dar el resultado o costo más bajo por campaña

Basado en el campo “**Cost per Result**” de una campaña como nuestra variable objetivo a predecir.

- **Cost per Result** = *The average cost per result from your ads.* -

Atributos: Para nuestro caso la edad

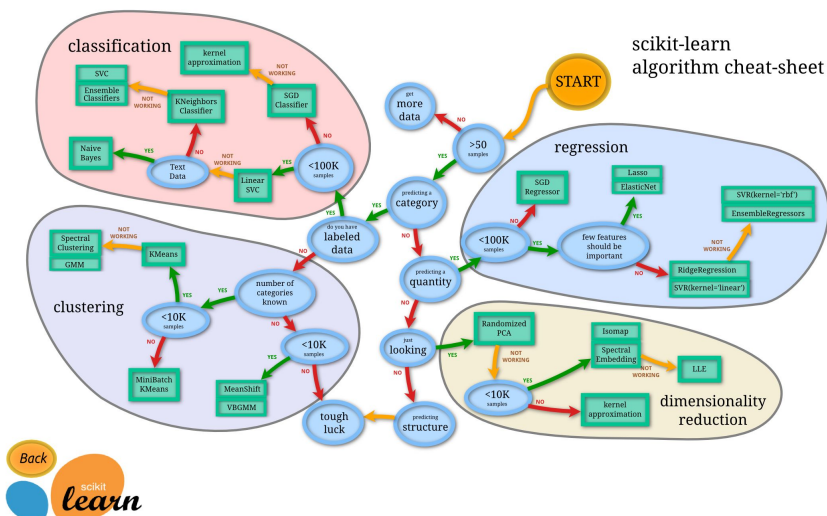
Solución propuesta: Modelo de regresión lineal (LASSO) y posteriormente un ejemplo con Random Forest

- Enfoque general, describir el problema predicción
- Enfoque a detalle, modelos usados,

Proyecto: Solución #1

Marco teórico

Regresión frente a clasificación



LASSO

- Predicting Quantity
- Samples 50-100,000
- Few Featurings are important

$$\min_w \frac{1}{2n_{\text{samples}}} \|Xw - y\|_2^2 + \alpha \|w\|_1$$

Contenido

Problemática

**Solución
Propuesta**

**¿Dónde
estamos?**

**Siguientes
pasos**

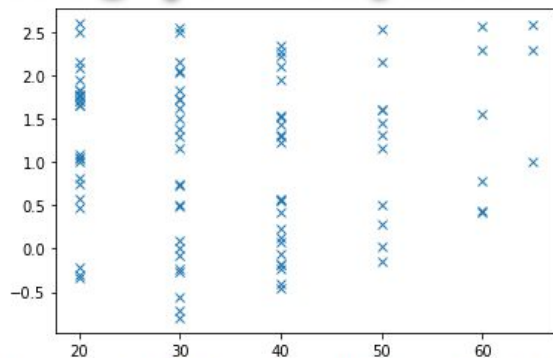
**¿Qué
aprendimos?**

¿Dónde estamos?

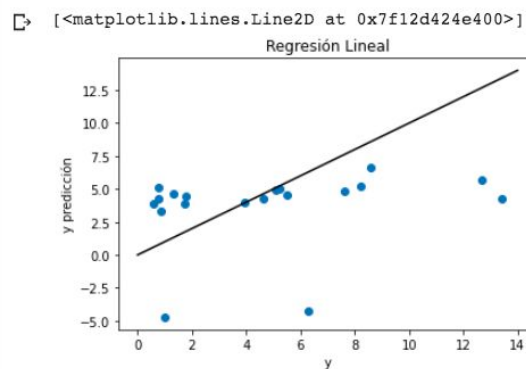


	Age	Goal	Impressions	Frequency	Reach	Results	Amount Spent (MXN)	Cost per Result
Age	1.000000	-0.051494	-0.060033	0.141627	-0.072817	0.014049	0.065579	0.061758
Goal	-0.051494	1.000000	-0.119884	-0.215846	-0.089495	0.163268	0.115432	0.813997
Impressions	-0.060033	-0.119884	1.000000	0.301487	0.947426	-0.102747	-0.090374	-0.300164
Frequency	0.141627	-0.215846	0.301487	1.000000	0.052662	-0.129093	-0.115971	-0.126214
Reach	-0.072817	-0.089495	0.947426	0.052662	1.000000	-0.098202	-0.082485	-0.256601
Results	0.014049	0.163268	-0.102747	-0.129093	-0.098202	1.000000	0.958186	0.175955
Amount Spent (MXN)	0.065579	0.115432	-0.090374	-0.115971	-0.082485	0.958186	1.000000	0.134756
Cost per Result	0.061758	0.813997	-0.300164	-0.126214	-0.256601	0.175955	0.134756	1.000000

$X = \text{Age} \mid Y = \text{Cost per result}$



```
plt.plot(df["Age"], np.log(df["Cost per Result"]), "x",
```



```
modelo.score(X_test, y_test)
```

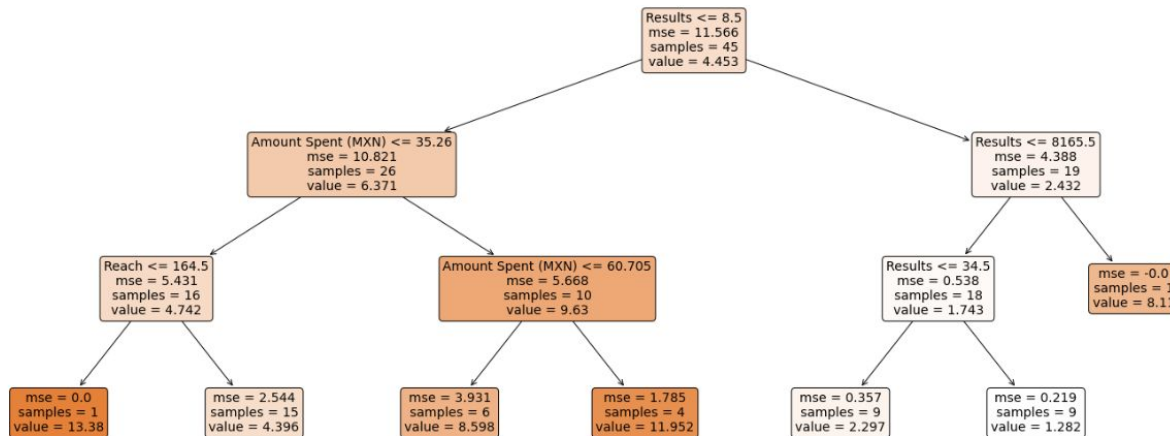


```
-0.30119451693475363
```

¿Hasta dónde llegamos?

Solución #2 - Modelo Random Forest

L J
C



Atributos para predecir el Cost per Result [Edad=30, Impresiones= 20000, Frecuencia=1, Alcance=80, Engadgment=6, Amount Spend=300]

```
[ ] modelo2.predict(np.reshape([30, 20000, 1, 80, 6, 300], (1,6)))
```

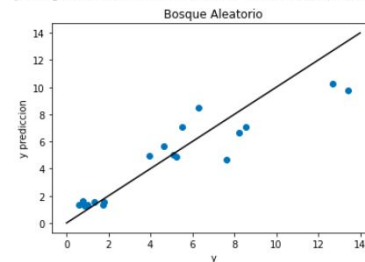
C array([10.15029414])



▶ modelo2.score(X_test, y_test)

▶ 0.8433980560772725

▶ [matplotlib.lines.Line2D at 0x7f12d2580668]



Contenido

Problemática

**Solución
Propuesta**

**¿Dónde
estamos?**

**Siguientes
pasos**

**¿Qué
aprendimos?**

Proyecto: Siguientes pasos

1. Obtener **más datos** de otras campañas de Facebook
2. Analizar los datos de **campañas de Google**
3. Aplicar **modelo** con los nuevos datos

Contenido

Problemática

**Solución
Propuesta**

**¿Dónde
estamos?**

**Siguientes
pasos**

**¿Qué
aprendimos?**

Lecciones Aprendidas

- La incorporación del **nivel Absolute Beginners** en Monterrey fue de mucha ayuda. Profesionistas que no son científicos de datos requieren guía para entender los conceptos.
- La guía/clases semanales del **mentor** marcaron la diferencia.
- Es importante dedicar tiempo al **autoestudio** y **ejercicios** conforme al plan.
- Es importante tomar la decisión de **presentar** o no un proyecto dos meses antes de terminar para alcanzar a **obtener los datos**.
- **¡Muchas gracias a Eduardo y Gerardo por organizar y a nuestro mentor estrella Eric por compartir con nosotros su experiencia de una manera tan práctica que a nuestro nivel pudiéramos entenderlo!**
- Definitivamente **estamos decididos a continuar en la siguiente generación** para seguir aprendiendo y **avanzando a los otros niveles**.

- Comprender la gestación
- Próximos Retos

Mayo 2020



Saturdays.AI

Saturdays.AI Monterrey
Grupo Absolute Beginners

Para todos los que hacen esto realidad

GRACIAS