# Nombre: Paulo Gonzalez

Objetivo:

* Consolidar los conocimientos adquiridos en clase para desarrollar simulaciones de eventos discretos.

Introducción:

El golpe económico de la crisis sanitaria del corona virus no va a ser cosa de semanas, sino de meses. Dentro de una de las etapas importantes posteriores a las elecciones presidenciales son el análisis y tendencia que tiene el actual presidente. Para ello se plantea realizar un sistema de regresión que permita identificar cual es la tendencia en base al manejo de las redes sociales (Twitter **y/o** Facebook) del presidente [1].

Las regresiones lineales pueden aprenden por sí mismos y en este caso obtener automáticamente esa “recta o ecuación” que buscamos con la tendencia de predicción. Para ello se puede midir el error con respecto a los puntos de entrada y el valor “Y” de salida real [3].

Enunciado:

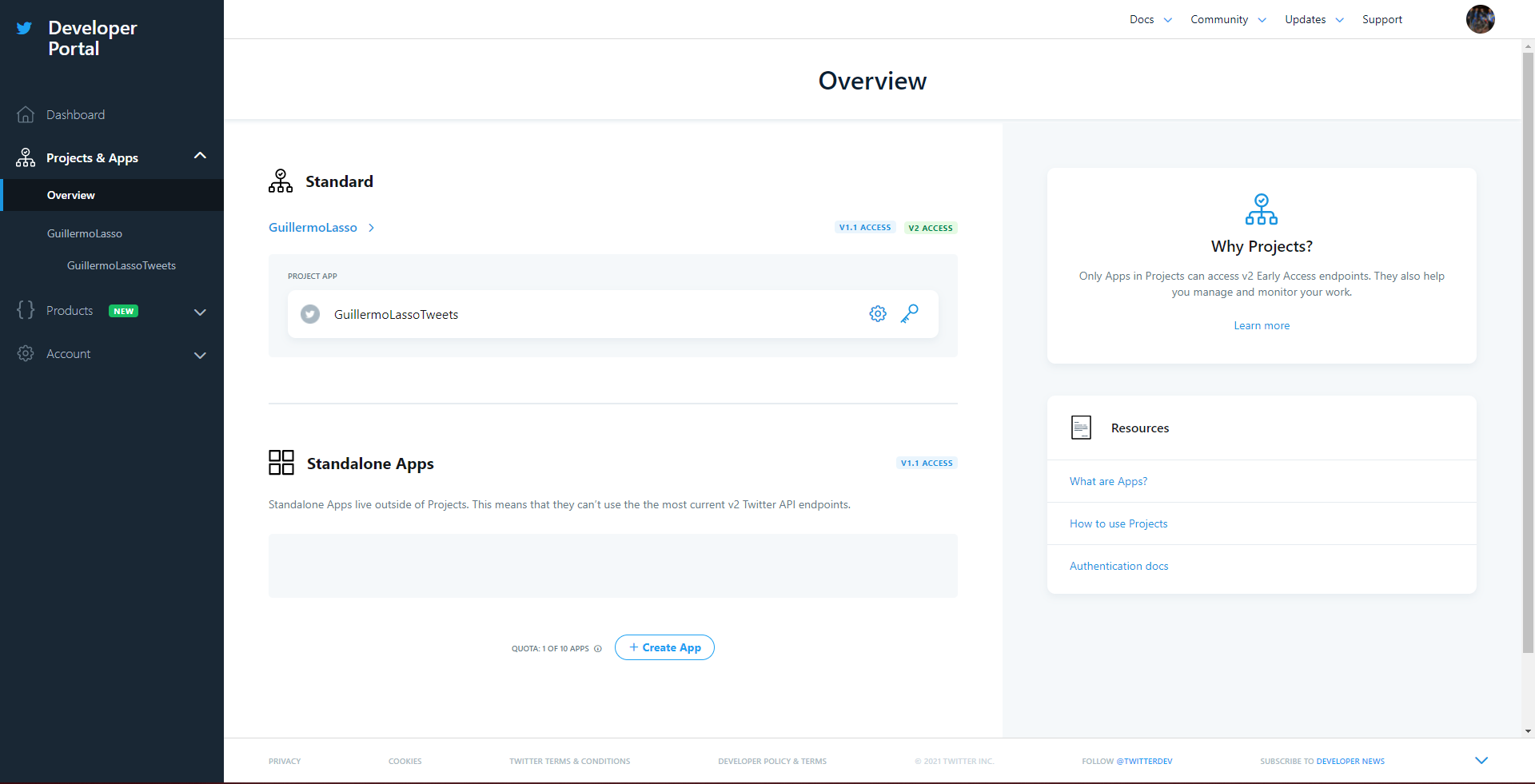
* Diseñe y desarrolle un modelo y/o script que permita simular el siguiente caso real:
  + Obtener datos de tendencia de twitter o facebook(crawler o webscraping), para ello se puede obtener a través del API [4].
* Title: Titulo del Post/Twitter
* Word count: la cantidad de palabras del artículo,
* # of Links: los enlaces externos que contiene,
* # of comments: cantidad de comentarios,
* # Shares: compartidos.
* **HashTag**
* Etc.

En base a ello, se pretende proponer y generar una predicción de cuántas veces será compartido un post/twitter utilizando regresión [2].

**DESARROLLO OBTENCION DE TWEETS (PUNTO 1)**

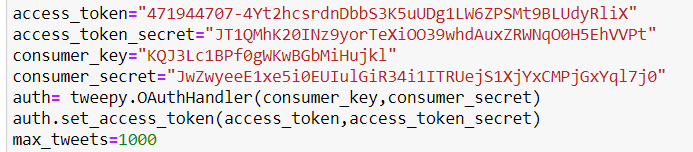
1. **Obtención de Tweets**

Para este punto se trabajo con la API que nos proporciona twitter; para ello generamos una cuenta de desarrollador en donde nos registraremos y generaremos una aplicación, la cual nos estrega las keys y el token para realizar el consumo de la API y obtener todos los tweets relacionados con Guillermo Lasso



Una vez que obtenemos esa información trabajaremos con la librería Tweepy que nos permite acceder y consumir la API de Twitter, con la cual generaremos una función para obtener los tweets de Guillermo Lasso y con ello generar un CSV que nos servirá para construir un modelo de regresión lineal y polinomial; cabe recalcar que Tweepy nos entrega información como un JSON, por lo cual lo almacenaremos en un archivo .TXT que luego convertiremos a CSV.

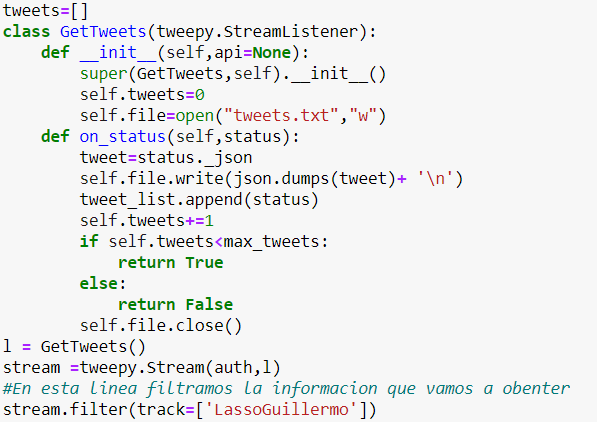




Debemos establecer nuestros tokens y key para acceder a la API, además estableceremos el máximo de tweets a ser consumidos.

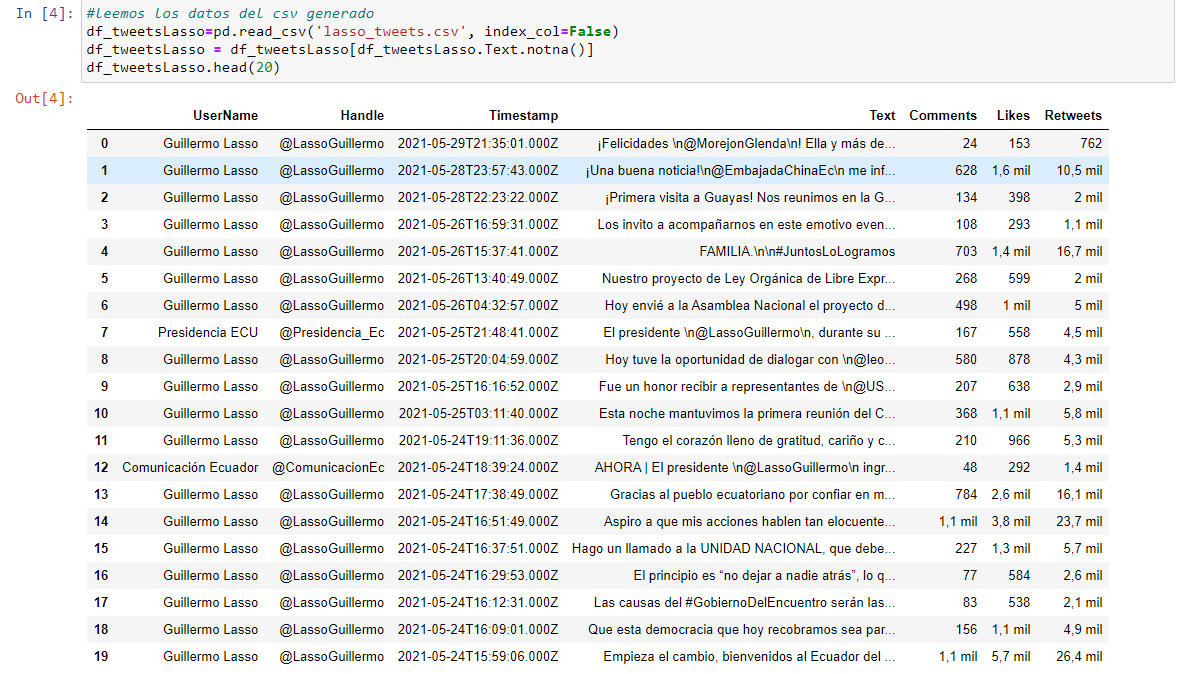
Ahora generamos un método en el cual obtendremos la información de los tweets además generaremos el CSV, esto ya que obtenemos un .txt que lo podemos transformar en un .csv, tweets[] será el arreglo que contemplará y almacenar cada tweet que este relacionado con Guillermo Lasso

Este método llamará la API de twitter, la cual se encargará de devolvernos información según el filtro que le hemos puesto, en este caso el método hará un llamado a la API y por medio de un for recorrerá y almacenar en un JSON toda la información relacionada con Guillermo Lasso.

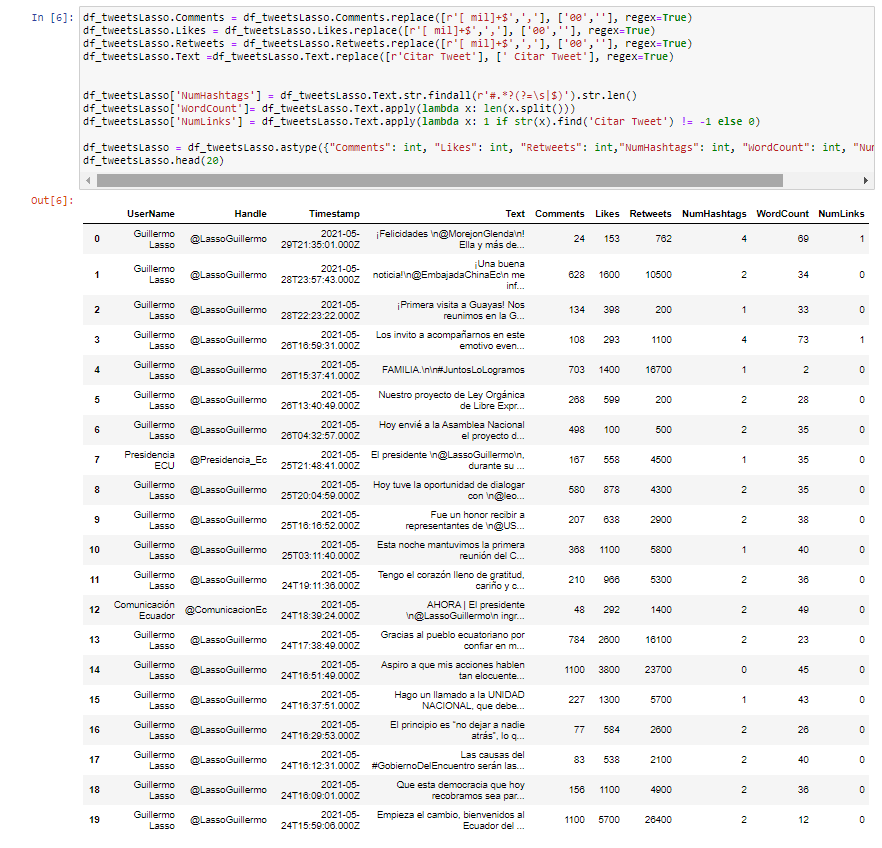


1. **Lectura y limpieza de datos**

Una vez obtenido el TXT procedemos a transformar a un CSV y sus datos en string ya que con ello podemos trabajar de una mejor manera.

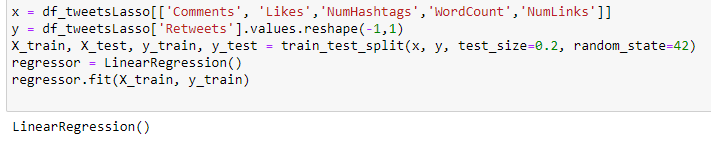


Ahora procedemos a hacer una limpieza de datos, ya que como podemos ver tenemos datos mezclados con números y letras, lo cual nos dará una complejidad al momento de generar el modelo y la predicción; además realizo cambio en nombre de las columnas para tener una mejor visibilidad y compresión de cada uno de los datos obtenidos.

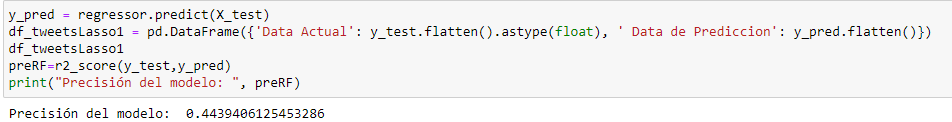


1. **Modelo de regresión Lineal**

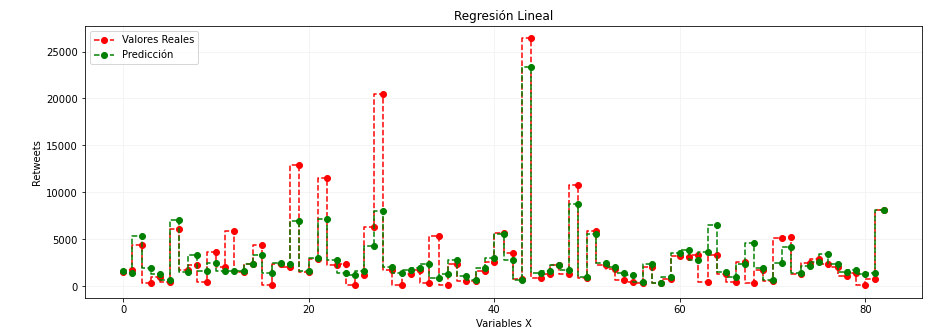
Para el modelo generamos un entrenamiento en el cual tomamos en cuenta lo que contiene un tweet contra los retweets generados, para ello usaremos un train\_test en donde estarán los datos y por medio de una librería (linearRegression) generar un valor de predicción.



Ahora ya generado nuestro entrenamiento vamos a generar un score para ver nuestro nivel de predicción.

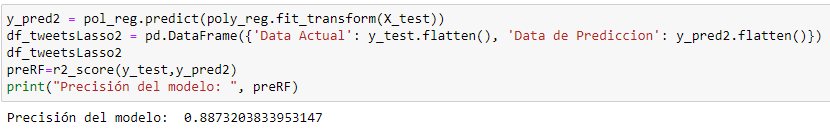


Ahora vemos nuestra grafica resultante.

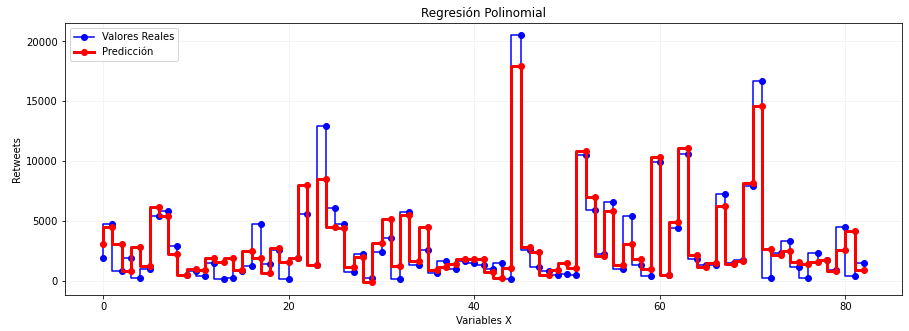


1. **Modelo de regresión Polinomial**

De igual manera seguimos los pasos anteriores por medio de un entramiento, la única diferencia es que trabajaremos con la librería PolynomialFeatures para nuestra regresión polinomial, dando como resultado de nuestro score



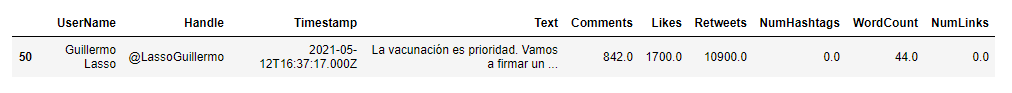
La grafica resultante es la siguiente



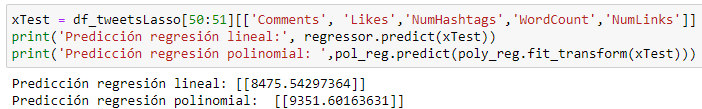
1. **Predicción de re-tweets**

Seleccionamos un tweet aleatorio a predecir sus posibles retweets, para ellos trabajaremos con un modelo de regresión lineal y polinomial.

El Tweet seleccionado es sobre la prioridad de vacunación.



Ahora generamos nuestro entrenamiento y trabajamos con las librerías mencionadas anteriormente; esto nos da como resultado:



* Posteriormente se debe seguir un proceso de vacunación en los recintos electorales que se describe a continuación (**Tomar el proceso para el proyecto final**):
  + Solo se va a tener en cuanta uno de los recintos electorales (investigar datos de cuantas personas asisten a votar).
  + Tomar los resultados de la regresión para la vacuna según la llegada.
  + Se tiene una promedio que el 80% de personas realizaran el proceso de vacunación dentro del Ecuador.
  + Dentro del procesos se tiene que alrededor del 5% - 10% no podrán vacunarse.
  + Las personas solo tiene un recinto electoral para realizar el proceso.
  + Las personas realizan la primera vacuna y 30 días después la segunda vacuna.
  + La persona se acerca a la mesa y hacen fila en caso de ser necesario para recibir la vacuna.
  + Realiza la vacunación en un tiempo aleatorio entre 5 a 10 minutos.
  + Debe esperar 20 minutos dentro del establecimiento para verificar que no tenga problemas de salud.
  + La persona recibe su certificado de vacunación y la fecha de la próxima vacuna entre 2 – 3 minutos.
  + La persona sale del recinto electoral.
  + Regresan para la próxima fecha y se repite el ciclo.
* El proceso de simulación desarrollado deberá considerar los siguientes aspectos:
  + Se debe establecer un modelo basado en modelos matemáticos para la predicción del numero de veces que se compartirá o la tendencia del presidente basada en redes sociales.
  + El programa deberá generar gráficas que indiquen la ecuación matemática de las tendencias .
  + Deben calcularse las siguientes métricas del sistema de simulación de eventos discretos :
    - Total de de personas que realizaron el proceso de vacunación.
    - Grafico del porcentaje de personas que no recibieron la vacuna.
    - El tiempo promedio de espera.

* **Fecha se presentación: 01/06/2021 – 23:55.**
* El informe en pdf (Archivo Notebook) se deberá subir al Git personal, todo el sistema debe ser basado en Python – Sympy - Open Source.

Opcional: Cualquier adición o librerías que no se revisaron en clases serán tomada como puntos adicionales al interciclo.

**Proyecto Final Presentación: 08/06/2021**

Referencias:

[1]: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6291769/>

[2]: <https://www.aprendemachinelearning.com/regresion-lineal-en-espanol-con-python/>

[3]: <https://eprints.ucm.es/48804/1/TFM%20Manuel%20Alejandro%20Rodriguez%20Santana.pdf>

[4]: <http://opac.pucv.cl/pucv_txt/txt-8000/UCC8094_01.pdf>