



**CENTRO DE CIENCIAS BÁSICAS**  
**DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN**  
**METAHEURÍSTICAS I**  
**7° "A"**

**PROYECTO FINAL: OPTIMIZAR EL TIEMPO DE PROCESAMIENTO GLOBAL EN LA  
DETECCIÓN DE PATRONES RELACIONADOS CON EL RACISMO EN REDES  
SOCIALES**

**Profesor: Francisco Javier Luna Rosas**

**Alumnos:**

**Almeida Ortega Andrea Melissa**  
**Espinoza Sánchez Joel Alejandro**  
**Flores Fernández Óscar Alonso**  
**Gómez Garza Dariana**  
**González Arenas Fernando Francisco**  
**Orocio García Hiram Efraín**

**Fecha de Entrega:** Aguascalientes, Ags., 26 de noviembre de 2021

# Índice

Introducción -----	1
Marco Teórico -----	3
Objetivos -----	4
Redes Sociales -----	5
Minería de la Web -----	10
Metodología -----	19
Computación Paralela -----	20
Aprendizaje Automático -----	53
Conclusiones -----	61
Referencias -----	65
Anexos -----	69

# Introducción

El racismo ha sido una tendencia de pensamiento que ha segregado a la población debido a la ideología que este trato a otras personas trae como consecuencia.

La ley de la “supervivencia del más apto” se formuló en la segunda mitad del siglo XIX. Pero la idea de que las relaciones sociales se reducen a una lucha permanente por sobrevivir surge mucho antes, de la mano de la ideología burguesa. Esto es manifiesto en los primeros tratados sobre el Estado, en particular en la obra de Thomas Hobbes y su visión de “la existencia humana como una bellum omnium contra omnes, una guerra de todos contra todos, que conduce a un estado de relaciones humanas de competitividad, desconfianza mutua y deseo de gloria” (Bohórquez, 2020).

La historia del racismo la comenta Tomás (2021) quien dice que la noción de “raza” toma un significado para el europeo al encontrarse con todo aquel no europeo durante la conquista. Es así como podemos encontrar que una definición de racismo es “la lectura y la atribución automática, prejuiciosa, de características intelectuales y morales que, de alguna forma, son inherentes a esos cuerpos” (Tomás, 2021).

Tomás describe también que el racismo fue un rasgo que se atribuyen a comienzos de la época moderna, sin embargo, él menciona que fue en los siglos XVIII y XIX cuando se empezó a considerar como teoría científica por Joseph Arthur de Gobineau.

Después de la Segunda Guerra Mundial, organizaciones como la UNESCO conceptualizaron el racismo principalmente como “una cuestión de actitudes individuales prejuiciosas y falta de educación” (Letin, 2004).

El hecho de que determinados “grupos” sean mucho más frecuente objeto de representaciones y estereotipos humillantes es asumido como algo dado, un dato social extremo cuyos fundamentos históricos no revisten interés analítico (Henriques, 2003).

Lejos de reducir el racismo a tendencias psicológicas históricas e internas de cada persona, estos enfoques están conceptualizados como si fuera un sistema de dominación y una cuestión de poder asimétrico. En los Estados Unidos, a finales de los sesenta, Ture y Hamilton cuñaron la expresión “racismo institucional” para señalar que el racismo no se manifiesta únicamente en los actos intencionales de los racistas “abiertos”, reproduciéndose mediante dinámicas sociales más profundas o incluso impersonales, produciendo hacer inferiores a las poblaciones no blancas en diferentes ámbitos y originándose “en las operaciones de fuerzas establecidas y respetadas socialmente” (Ture y Hamilton, 1992).

En los años 2010, 2011 y 2013 se hizo la publicación del “Panel sobre la discriminación por origen racial o étnico”, un estudio sobre la percepción de la discriminación por parte de las potenciales víctimas. El panel evidenciaba que los “grupos” afrodescendiente, Rom/gitano, marroquí y andino eran los más afectados por el racismo y señalaba como principales ámbitos de discriminación como la vivienda, el trato policial, el empleo, el acceso a espacios de ocio y servicios públicos (Consejo 2011, 2012,2014).

# Marco Teórico

## Caso de estudio: El racismo estudiado en el idioma inglés

El 1° de octubre del 2020, bajo 128 resoluciones, declaraciones o en general, contenido legislativo relacionado al racismo, el gobierno de los Estados Unidos declaró al racismo como una crisis de salud pública dado como dato por Dara (2021) quien provee en su artículo una tabla que muestra detalladamente con los datos de cada estado de la nación.

Las políticas tomadas debido a las acciones legislativas tomadas y relatadas por el artículo de Dara llevan de manera consecuente un posible primer paso – según Dara – de modo que se archive la inequidad por el racismo, pues “en muchas ocasiones, estas razones se entremezclan, incluso de forma inconsciente, y dan lugar a actitudes racistas que vulneran los derechos humanos.” (Quintero, 2021).

Una extensión de estudio para analizar el racismo dentro de esta situación será poder determinar perfiles y emociones dentro de la expresión social del internet y así emitir una hipótesis con base en la observación de los sentimientos en ellos.

Esta forma de examinar se basa en que “cada persona tiene la capacidad de realizar acciones concretas para evitar la conversión de Internet en un espacio lleno de odio. La aportación de datos objetivos y actualizados, así como las estrategias emocionales, son la clave para desmontar falsos rumores. Recientemente están apareciendo diversas guías que aportan herramientas prácticas para responder al discurso de odio” (Quintero, 2021). Patrones que por medio de la presente investigación se piensan encontrar.

El rumbo de este proyecto presenta una investigación con relación al racismo y su relación con las redes sociales; después se realizará clasificación y aprendizaje automático y luego la optimización relacionada al Big Data del proyecto.

# Objetivos

## Objetivo general

El objetivo general del presente proyecto es optimizar el tiempo de procesamiento global en la detección de patrones relacionados con el racismo en redes sociales.

## Objetivos específicos

Los objetivos específicos que se planteó el equipo son los siguientes:

- Crear un modelo de reconocimiento de patrones relacionados con el racismo.
- Simular una máquina de aprendizaje en paralelo para optimizar tiempo.

# Redes Sociales

## Colección de datos

El estudio se hará por medio de redes sociales, en concreto, de la red social Twitter, que propiciará al programa un conjunto muy grande de pensamientos, opiniones cortas y expresiones – coloquialmente conocidas como tweets – con el que se analizarán los rasgos más comunes en la comunicación textual para intuir si éstos son maliciosos o no.

Dado que se realizará – en pocas palabras – un análisis de sentimientos con la herramienta Python, éste requiere un vocabulario inicial para examinar los tweets que se extraerán de la red social. A continuación se presenta el vocabulario a utilizar en esta fase de la investigación.

Vocabulario	
Inglés	
Literatura	Contenido
UC Santa Barbara linguistics expert (April, 2021)	"sold down the river", "Master bedrooms", "Blacklists", "whitelists", "Cakewalk", "Lynch mob", "Uppity", "Blackball, black mark",
Elizabeth Pryor, an associate professor of history at Smith College. (April, 2021)	Slave, Master
Donald Trump (October, 2020)	"Lynching in America"
2016, Naomi Campbell in her autobiography, <i>Naomi Campbell</i> .	"I didn't let it rattle me. From attending auditions and performing at an early age, I understood what it meant to be black. You had to put in the extra effort. You had to be twice as good."
Christina Sterbenz and Dominic-Madori Davis (June, 2020)	"The itis", "Peanut gallery", "Gyp", "Paddy wagons", "Bugger", "Hooligan", "Eskimo", "Eenie meenie miney moe", "Cretins",
Alan Connor, author of "The Crossword Century" (July, 2020)	"Open de kimono", "Fuzzy wuzzy", "Off the reservation", "Indian style", "Mumbo jumbo",
Christopher Cicchiello, (July, 2020)	"Freeholder", "Jimmies", "Black sheep"
H.L. Mencken (1919)	"Negroes", "Skunk", "Black bird", "Spade lips"
Propuesto en esta investigación	Karen, "Go back to your country!", "This is the United States of America!", "What



---

is this, Nazi Germany?", "English only",  
"Chinese guy", "racism"

---

Vocabulario	
Español	
Literatura	Contenido
Patricio Solís, investigador del Colegio de México (Colmex).	"La discriminación étnico-racial es estructural porque se funda en un orden social y en una relación de poder que tienen antecedentes históricos en el país desde hace varios siglos y se reproduce de manera permanente en la sociedad, por lo que también sus efectos son estructurales"
Eugenio Derbéz, (Noviembre, 2016)	"Desde que hablas español ya te ven diferente, desde que llegas, a mi me pasaron a un cuartito muchas veces nada más por el color de piel o porque te notan el acento"
Jorge Ramos, después de su careo con Donald Trump en 2015 en Iowa.	"[...] uno de sus seguidores a las afueras de la conferencia de prensa de donde me echaron, me dice: 'Lárgate de mi país'".
César Évora	"[...] creen que los latinos son en su mayoría de piel morena, aunque cuando se encuentran con alguien que no tiene esos rasgos, les cuesta entender que son mexicanos, cubanos, etc."
Maluma, Al punto (2016)	"Yo no vivo en Estados Unidos, pero sí tengo voz me gustaría lanzar un

	mensaje que diga claramente que nos unamos, todos, para intentar luchar por nuestros derechos".
Propuesto en esta investigación	"el mantecas", "Peruano", "Venezolano" "Gringo" "Güerito de rancho", "Chino cochino", "Whitexicans", "Shrekxicans", "Buchona", "El aladín", "Vas a mejorar la raza", "Negro", "El mofles", "Zamorita", "Negrita cucurumbé", "Frijol negro", "Come palomas", "Trabajo como negro para vivir como blanco", "Nunca faltó un prieto en el arroz", "Indio", "Chacha", "Gata", "Traes el nopal en la cara", "Café con leche"

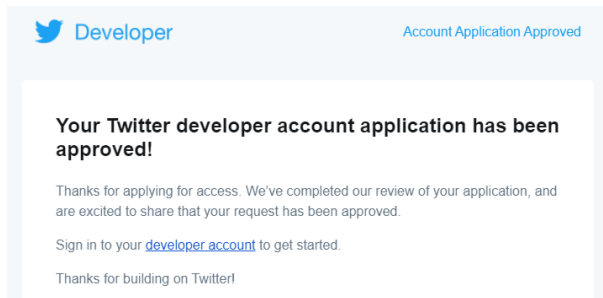
# Minería de la Web

Como parte de la minería del contenido de la web, se realizará aprendizaje supervisado, concretamente realizando clasificación sobre textos.

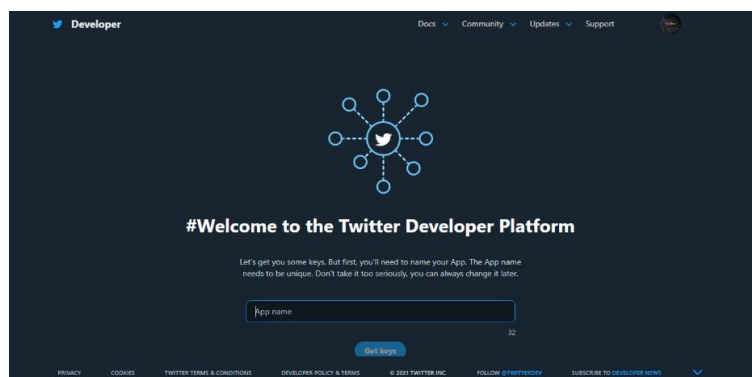
En principio se instaló la librería a usar en la tecnología de Python que será Tweepy:

```
base) C:\Users\alexe>pip install tweepy
Collecting tweepy
  Downloading tweepy-3.10.0-py2.py3-none-any.whl (30 kB)
Requirement already satisfied: requests[socks]>=2.11.1 in c:\programdata\anaconda3\lib\site-packages (from tweepy) (2.24.0)
Requirement already satisfied: requests-oauthlib>=0.7.0 in c:\programdata\anaconda3\lib\site-packages (from tweepy) (1.3.0)
Requirement already satisfied: six>=1.10.0 in c:\programdata\anaconda3\lib\site-packages (from tweepy) (1.15.0)
Requirement already satisfied: chardet<4,>=3.0.2 in c:\programdata\anaconda3\lib\site-packages (from requests[socks]>=2.11.1->tweepy) (3.0.4)
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in c:\programdata\anaconda3\lib\site-packages (from requests[socks]>=2.11.1->tweepy) (1.25.11)
Requirement already satisfied: certifi>=2017.4.17 in c:\programdata\anaconda3\lib\site-packages (from requests[socks]>=2.11.1->tweepy) (2020.6.20)
Requirement already satisfied: idna<3,>=2.5 in c:\programdata\anaconda3\lib\site-packages (from requests[socks]>=2.11.1->tweepy) (2.10)
Requirement already satisfied: PySocks!=1.5.7,>=1.5.6; extra == "socks" in c:\programdata\anaconda3\lib\site-packages (from requests[socks]>=2.11.1->tweepy) (1.7.1)
Requirement already satisfied: oauthlib>=3.0.0 in c:\programdata\anaconda3\lib\site-packages (from requests-oauthlib>=0.7.0->tweepy) (3.1.0)
Installing collected packages: tweepy
Successfully installed tweepy-3.10.0
```

Posteriormente se realizó el procedimiento necesario para poder acceder a la cuenta de desarrollador de Twitter



Posteriormente se realizó el proyecto sobre la plataforma de desarrollador de Twitter como se muestra a continuación:



Previo a realizar una implementación en código, se investigaron en su totalidad los atributos de un tweet que se extraen, a saber, los siguientes son las características guardadas por tweet en la plataforma:

```
['_class_', '__delattr__', '__dict__', '__dir__', '__doc__', '__eq__', '__format__', '__ge__', '__getattr__', '__getstate__', '__gt__', '__hash__', '__init__', '__init_subclass__', '__le__', '__lt__', '__module__', '__ne__', '__new__', '__reduce__', '__reduce_ex__', '__repr__', '__setattr__', '__sizeof__', '__str__', '__subclasshook__', '__weakref__', '_api', '_json', 'author', 'contributors', 'coordinates', 'created_at', 'destroy', 'display_text_range', 'entities', 'favorite', 'favorite_count', 'favorited', 'full_text', 'geo', 'id', 'id_str', 'in_reply_to_screen_name', 'in_reply_to_status_id', 'in_reply_to_status_id_str', 'in_reply_to_user_id', 'in_reply_to_user_id_str', 'is_quote_status', 'lang', 'parse', 'parse_list', 'place', 'retweet', 'retweet_count', 'retweeted', 'retweets', 'source', 'source_url', 'truncated', 'user']
```

Se realizó primeramente un código prueba para comprobar los datos recopilados por cada atributo (véase anexo 1):

```
aut = tweepy.OAuthHandler(k1, jorge)
aut.set_access_token(k3, k4)
api = tweepy.API(aut)
numt = 750
autor = []
contri = []
coord = []
cread = []
destroy = []
disp_text = []
enti = []
fav = []
fav_count = []
favd = []
ftxt = []
geo = []
iid = []
id_str = []
inscream = []
inrepstat = []
inrapstatstr = []
inrusid = []
inruseridstr = []
iscuotstat = []
lang = []
parse = []
parse_list = []
plays = []
retwitt = []
retwitt_count = []
retwitter = []
retwitts = []
source = []
source_url = []
truncater = []
user = []
```

```

twit = tweepy.Cursor(api.search, q="nigger", tweet_mode="extended").items(numt)
for i in twit:
    autor.append(i.author)
    contri.append(i.contributors)
    coord.append(i.coordinates)
    cread.append(i.created_at)
    destroy.append(i.destroy)
    disp_text.append(i.display_text_range)
    enti.append(i.entities)
    fav.append(i.favorite)
    fav_count.append(i.favorite_count)
    favd.append(i.favorited)
    ftxt.append(i.full_text)
    geo.append(i.geo)
    iid.append(i.id)
    id_str.append(i.id_str)
    inscream.append(i.in_reply_to_screen_name)
    inrepsat.append(i.in_reply_to_status_id)
    inrapstatstr.append(i.in_reply_to_status_id_str)
    inrusid.append(i.in_reply_to_user_id)
    inruseridstr.append(i.in_reply_to_user_id_str)
    iscutstat.append(i.is_quote_status)
    lang.append(i.lang)
    parse.append(i.parse)
    parse_list.append(i.parse_list)
    plays.append(i.place)
    retwitt.append(i.retweet)
    retwitt_count.append(i.retweet_count)
    retwitter.append(i.retweeted)
    retwitts.append(i.retweets)
    source.append(i.source)
    source_ulr.append(i.source_url)
    truncater.append(i.truncated)
    user.append(i.user)

df = pd.DataFrame({'autor':autor,'contri':contri,'coord':coord,'cread':cread, 'destroy':destroy,'disp_text':disp_text,
df.to_csv()

```

Se obtuvo un archivo CSV muy extenso que se analizaron los atributos y cuáles serían útiles para el trabajo a realizar. El archivo se muestra a continuación:

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	autor	contri	coord	cread	destroy	disp_text	enti	fav	fav_count	favd	ftxt	geo	iid	id_str	inscream	inrepsat	inrapstatstr	inrusid	inruseridstr	iscutstat	lang
2	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
3	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
4	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
5	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
6	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
7	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
8	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
9	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
10	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
11	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
12	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
13	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
14	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
15	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
16	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
17	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
18	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
19	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
20	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
21	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
22	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
23	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
24	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
25	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
26	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
27	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
28	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
29	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
30	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
31	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
32	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
33	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
34	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
35	User_api=tweepy.api.API object at 0: #####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####

Se analizaron cada uno de los atributos en este archivo. En este análisis se concluyó que muchos de ellos contenían información repetida, inaccesible o simplemente, irrelevante. Se eliminó toda esta información para la siguiente etapa de código.

A partir de este hecho, se inició la implementación formal en código de lenguaje Python con el siguiente código (véase anexo 2):

```
### Uso de librerías
import pandas as pd
import tweepy

### Lectura de llaves
stream = open("keys.py")
read_file = stream.read()
exec(read_file)

### Conexión con la API
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)

### Definición de valores
num_tweets = 900
created_at = []
description = []
favorite_count = []
full_text = []
identifier = []
in_reply_to_screen_name = []
lang = []
location = []
name = []
screen_name = []
tweet = tweepy.Cursor(api.search, q="nigger", tweet_mode="extended").items(num_tweets)

























































### Extracción de tweets
for i in tweet:
    created_at.append(i.created_at)
    description.append(i.author.description)
    favorite_count.append(i.favorite_count)
    full_text.append(i.full_text)
    identifier.append(i.id)
    lang.append(i.lang)
    location.append(i.author.location)
    in_reply_to_screen_name.append(i.in_reply_to_screen_name)
```

Generando una base de datos del siguiente tipo:

[illegible]











































































El equipo realizó extracciones de 900 tweets por archivo y para tratar de realizar la simulación de Big Data, se trataron de recolectar la mayor cantidad de tweets.

Melissa logró 430 archivos de tweets para el proyecto:











































































 MelScrap 403.csv		05/10/2021 09:50 a. m.	Archivo de valores...	225 KB
 MelScrap 404.csv		05/10/2021 09:50 a. m.	Archivo de valores...	226 KB
 MelScrap 405.csv		05/10/2021 09:50 a. m.	Archivo de valores...	225 KB
 MelScrap 406.csv		05/10/2021 09:50 a. m.	Archivo de valores...	224 KB
 MelScrap 407.csv		05/10/2021 09:50 a. m.	Archivo de valores...	224 KB
 MelScrap 408.csv		05/10/2021 09:50 a. m.	Archivo de valores...	224 KB
 MelScrap 409.csv		05/10/2021 09:50 a. m.	Archivo de valores...	224 KB
 MelScrap 410.csv		05/10/2021 09:50 a. m.	Archivo de valores...	224 KB
 MelScrap 411.csv		05/10/2021 09:50 a. m.	Archivo de valores...	225 KB
 MelScrap 412.csv		05/10/2021 09:50 a. m.	Archivo de valores...	225 KB
 MelScrap 413.csv		05/10/2021 09:50 a. m.	Archivo de valores...	224 KB
 MelScrap 414.csv		05/10/2021 09:50 a. m.	Archivo de valores...	1 KB
 MelScrap 415.csv		05/10/2021 09:50 a. m.	Archivo de valores...	2 KB
 MelScrap 416.csv		05/10/2021 09:50 a. m.	Archivo de valores...	4 KB
 MelScrap 417.csv		05/10/2021 09:50 a. m.	Archivo de valores...	6 KB
 MelScrap 418.csv		05/10/2021 09:50 a. m.	Archivo de valores...	9 KB
 MelScrap 419.csv		05/10/2021 09:50 a. m.	Archivo de valores...	12 KB
 MelScrap 420.csv		05/10/2021 09:50 a. m.	Archivo de valores...	14 KB
 MelScrap 421.csv		05/10/2021 09:50 a. m.	Archivo de valores...	17 KB
 MelScrap 422.csv		05/10/2021 09:50 a. m.	Archivo de valores...	18 KB
 MelScrap 423.csv		05/10/2021 09:50 a. m.	Archivo de valores...	21 KB
 MelScrap 424.csv		05/10/2021 09:50 a. m.	Archivo de valores...	23 KB
 MelScrap 425.csv		05/10/2021 09:50 a. m.	Archivo de valores...	26 KB
 MelScrap 426.csv		05/10/2021 09:50 a. m.	Archivo de valores...	31 KB
 MelScrap 427.csv		05/10/2021 09:50 a. m.	Archivo de valores...	35 KB
 MelScrap 428.csv		05/10/2021 09:50 a. m.	Archivo de valores...	233 KB
 MelScrap 429.csv		05/10/2021 09:50 a. m.	Archivo de valores...	233 KB
 MelScrap 430.csv		05/10/2021 09:50 a. m.	Archivo de valores...	233 KB













































































Joel realizó 1,290 archivos para el proyecto:

 JoulScrap 1254.csv		30/09/2021 07:57 p. m.	Archivo de valores...	233 KB
 JoulScrap 1255.csv		30/09/2021 08:12 p. m.	Archivo de valores...	232 KB
 JoulScrap 1256.csv		30/09/2021 08:27 p. m.	Archivo de valores...	232 KB
 JoulScrap 1257.csv		30/09/2021 08:42 p. m.	Archivo de valores...	233 KB
 JoulScrap 1258.csv		01/10/2021 10:47 a. m.	Archivo de valores...	236 KB
 JoulScrap 1259.csv		01/10/2021 11:02 a. m.	Archivo de valores...	236 KB
 JoulScrap 1260.csv		01/10/2021 11:17 a. m.	Archivo de valores...	236 KB
 JoulScrap 1261.csv		01/10/2021 11:32 a. m.	Archivo de valores...	236 KB
 JoulScrap 1262.csv		01/10/2021 11:47 a. m.	Archivo de valores...	237 KB
 JoulScrap 1263.csv		01/10/2021 12:02 p. m.	Archivo de valores...	237 KB
 JoulScrap 1264.csv		01/10/2021 12:17 p. m.	Archivo de valores...	237 KB
 JoulScrap 1265.csv		01/10/2021 12:32 p. m.	Archivo de valores...	237 KB
 JoulScrap 1266.csv		01/10/2021 12:47 p. m.	Archivo de valores...	238 KB
 JoulScrap 1267.csv		01/10/2021 01:02 p. m.	Archivo de valores...	238 KB
 JoulScrap 1268.csv		01/10/2021 01:17 p. m.	Archivo de valores...	239 KB
 JoulScrap 1269.csv		01/10/2021 01:32 p. m.	Archivo de valores...	238 KB
 JoulScrap 1270.csv		01/10/2021 01:47 p. m.	Archivo de valores...	238 KB
 JoulScrap 1271.csv		01/10/2021 02:02 p. m.	Archivo de valores...	237 KB
 JoulScrap 1272.csv		01/10/2021 02:17 p. m.	Archivo de valores...	238 KB
 JoulScrap 1273.csv		01/10/2021 02:32 p. m.	Archivo de valores...	236 KB
 JoulScrap 1274.csv		01/10/2021 02:47 p. m.	Archivo de valores...	236 KB
 JoulScrap 1275.csv		01/10/2021 03:02 p. m.	Archivo de valores...	236 KB
 JoulScrap 1276.csv		01/10/2021 03:17 p. m.	Archivo de valores...	236 KB
 JoulScrap 1277.csv		01/10/2021 03:32 p. m.	Archivo de valores...	237 KB
 JoulScrap 1278.csv		01/10/2021 03:47 p. m.	Archivo de valores...	237 KB
 JoulScrap 1279.csv		01/10/2021 04:02 p. m.	Archivo de valores...	239 KB
 JoulScrap 1280.csv		01/10/2021 04:17 p. m.	Archivo de valores...	239 KB
 JoulScrap 1281.csv		01/10/2021 04:32 p. m.	Archivo de valores...	239 KB
 JoulScrap 1282.csv		01/10/2021 04:47 p. m.	Archivo de valores...	239 KB
 JoulScrap 1283.csv		01/10/2021 05:02 p. m.	Archivo de valores...	240 KB
 JoulScrap 1284.csv		01/10/2021 05:17 p. m.	Archivo de valores...	240 KB
 JoulScrap 1285.csv		03/10/2021 11:30 a. m.	Archivo de valores...	212 KB
 JoulScrap 1286.csv		03/10/2021 11:45 a. m.	Archivo de valores...	212 KB
 JoulScrap 1287.csv		03/10/2021 12:00 p. m.	Archivo de valores...	211 KB
 JoulScrap 1288.csv		03/10/2021 12:15 p. m.	Archivo de valores...	212 KB
 JoulScrap 1289.csv		03/10/2021 12:30 p. m.	Archivo de valores...	211 KB
 JoulScrap 1290.csv		03/10/2021 12:45 p. m.	Archivo de valores...	214 KB

Óscar realizó 2,038 extracciones:











































































	OscarScrap 2002.csv		05/10/2021 11:52 a. m.	Archivo de valores...	234 KB
	OscarScrap 2003.csv		05/10/2021 11:52 a. m.	Archivo de valores...	234 KB
	OscarScrap 2004.csv		05/10/2021 11:52 a. m.	Archivo de valores...	234 KB
	OscarScrap 2005.csv		05/10/2021 11:52 a. m.	Archivo de valores...	234 KB
	OscarScrap 2006.csv		05/10/2021 11:52 a. m.	Archivo de valores...	234 KB
	OscarScrap 2007.csv		05/10/2021 11:52 a. m.	Archivo de valores...	234 KB
	OscarScrap 2008.csv		05/10/2021 11:52 a. m.	Archivo de valores...	234 KB
	OscarScrap 2009.csv		05/10/2021 11:52 a. m.	Archivo de valores...	234 KB
	OscarScrap 2010.csv		05/10/2021 11:52 a. m.	Archivo de valores...	233 KB
	OscarScrap 2011.csv		05/10/2021 11:52 a. m.	Archivo de valores...	233 KB
	OscarScrap 2012.csv		05/10/2021 11:52 a. m.	Archivo de valores...	234 KB
	OscarScrap 2013.csv		05/10/2021 11:52 a. m.	Archivo de valores...	234 KB
	OscarScrap 2014.csv		05/10/2021 11:52 a. m.	Archivo de valores...	233 KB
	OscarScrap 2015.csv		05/10/2021 11:52 a. m.	Archivo de valores...	234 KB
	OscarScrap 2016.csv		05/10/2021 11:52 a. m.	Archivo de valores...	234 KB
	OscarScrap 2017.csv		05/10/2021 11:52 a. m.	Archivo de valores...	234 KB
	OscarScrap 2018.csv		05/10/2021 11:52 a. m.	Archivo de valores...	232 KB
	OscarScrap 2019.csv		05/10/2021 11:52 a. m.	Archivo de valores...	233 KB
	OscarScrap 2020.csv		05/10/2021 11:52 a. m.	Archivo de valores...	234 KB
	OscarScrap 2021.csv		05/10/2021 11:52 a. m.	Archivo de valores...	234 KB
	OscarScrap 2022.csv		05/10/2021 11:52 a. m.	Archivo de valores...	235 KB
	OscarScrap 2023.csv		05/10/2021 11:52 a. m.	Archivo de valores...	235 KB
	OscarScrap 2024.csv		05/10/2021 11:52 a. m.	Archivo de valores...	234 KB
	OscarScrap 2025.csv		05/10/2021 11:52 a. m.	Archivo de valores...	233 KB
	OscarScrap 2026.csv		05/10/2021 11:52 a. m.	Archivo de valores...	233 KB
	OscarScrap 2027.csv		05/10/2021 11:52 a. m.	Archivo de valores...	232 KB
	OscarScrap 2028.csv		05/10/2021 11:52 a. m.	Archivo de valores...	224 KB
	OscarScrap 2029.csv		05/10/2021 11:52 a. m.	Archivo de valores...	225 KB
	OscarScrap 2030.csv		05/10/2021 11:52 a. m.	Archivo de valores...	224 KB
	OscarScrap 2031.csv		05/10/2021 11:52 a. m.	Archivo de valores...	224 KB
	OscarScrap 2032.csv		05/10/2021 11:52 a. m.	Archivo de valores...	225 KB
	OscarScrap 2033.csv		05/10/2021 11:52 a. m.	Archivo de valores...	223 KB
	OscarScrap 2034.csv		05/10/2021 11:52 a. m.	Archivo de valores...	221 KB
	OscarScrap 2035.csv		05/10/2021 11:52 a. m.	Archivo de valores...	220 KB
	OscarScrap 2036.csv		05/10/2021 11:52 a. m.	Archivo de valores...	220 KB
	OscarScrap 2037.csv		05/10/2021 11:52 a. m.	Archivo de valores...	222 KB
	OscarScrap 2038.csv		05/10/2021 11:52 a. m.	Archivo de valores...	222 KB

Dariana ha realizado 209 extracciones:

 DariScrap 173.csv		05/10/2021 09:28 a. m.	Archivo de valores...	287 KB
 DariScrap 174.csv		05/10/2021 09:28 a. m.	Archivo de valores...	295 KB
 DariScrap 175.csv		05/10/2021 09:29 a. m.	Archivo de valores...	295 KB
 DariScrap 176.csv		05/10/2021 09:29 a. m.	Archivo de valores...	293 KB
 DariScrap 177.csv		05/10/2021 09:29 a. m.	Archivo de valores...	290 KB
 DariScrap 178.csv		05/10/2021 09:29 a. m.	Archivo de valores...	295 KB
 DariScrap 179.csv		05/10/2021 09:29 a. m.	Archivo de valores...	291 KB
 DariScrap 180.csv		05/10/2021 09:29 a. m.	Archivo de valores...	288 KB
 DariScrap 181.csv		05/10/2021 09:29 a. m.	Archivo de valores...	292 KB
 DariScrap 182.csv		05/10/2021 09:29 a. m.	Archivo de valores...	291 KB
 DariScrap 183.csv		05/10/2021 09:29 a. m.	Archivo de valores...	294 KB
 DariScrap 184.csv		05/10/2021 09:29 a. m.	Archivo de valores...	285 KB
 DariScrap 185.csv		05/10/2021 09:29 a. m.	Archivo de valores...	291 KB
 DariScrap 186.csv		05/10/2021 09:29 a. m.	Archivo de valores...	294 KB
 DariScrap 187.csv		05/10/2021 09:29 a. m.	Archivo de valores...	288 KB
 DariScrap 188.csv		05/10/2021 09:29 a. m.	Archivo de valores...	285 KB
 DariScrap 189.csv		05/10/2021 09:29 a. m.	Archivo de valores...	282 KB
 DariScrap 190.csv		05/10/2021 09:29 a. m.	Archivo de valores...	285 KB
 DariScrap 191.csv		05/10/2021 09:29 a. m.	Archivo de valores...	289 KB
 DariScrap 192.csv		05/10/2021 09:29 a. m.	Archivo de valores...	287 KB
 DariScrap 193.csv		05/10/2021 09:29 a. m.	Archivo de valores...	295 KB
 DariScrap 194.csv		05/10/2021 09:29 a. m.	Archivo de valores...	296 KB
 DariScrap 195.csv		05/10/2021 09:29 a. m.	Archivo de valores...	292 KB
 DariScrap 196.csv		05/10/2021 09:29 a. m.	Archivo de valores...	298 KB
 DariScrap 197.csv		05/10/2021 09:29 a. m.	Archivo de valores...	295 KB
 DariScrap 198.csv		05/10/2021 09:29 a. m.	Archivo de valores...	292 KB
 DariScrap 199.csv		05/10/2021 09:29 a. m.	Archivo de valores...	293 KB
 DariScrap 200.csv		05/10/2021 09:29 a. m.	Archivo de valores...	288 KB
 DariScrap 201.csv		05/10/2021 09:29 a. m.	Archivo de valores...	287 KB
 DariScrap 202.csv		05/10/2021 09:29 a. m.	Archivo de valores...	290 KB
 DariScrap 203.csv		05/10/2021 09:29 a. m.	Archivo de valores...	285 KB
 DariScrap 204.csv		05/10/2021 09:29 a. m.	Archivo de valores...	286 KB
 DariScrap 205.csv		05/10/2021 09:29 a. m.	Archivo de valores...	290 KB
 DariScrap 206.csv		05/10/2021 09:29 a. m.	Archivo de valores...	289 KB
 DariScrap 207.csv		05/10/2021 09:29 a. m.	Archivo de valores...	288 KB
 DariScrap 208.csv		05/10/2021 09:29 a. m.	Archivo de valores...	293 KB
 DariScrap 209.csv		05/10/2021 09:29 a. m.	Archivo de valores...	297 KB

Fernando aportó con 178 tweets.

Hiram aportó con 1,485 tweets:

 EFraScrap 1449.csv		05/10/2021 10:37 a. m.	Archivo de valores...	239 KB
 EFraScrap 1450.csv		05/10/2021 10:37 a. m.	Archivo de valores...	239 KB
 EFraScrap 1451.csv		05/10/2021 10:37 a. m.	Archivo de valores...	239 KB
 EFraScrap 1452.csv		05/10/2021 10:37 a. m.	Archivo de valores...	239 KB
 EFraScrap 1453.csv		05/10/2021 10:37 a. m.	Archivo de valores...	239 KB
 EFraScrap 1454.csv		05/10/2021 10:37 a. m.	Archivo de valores...	234 KB
 EFraScrap 1455.csv		05/10/2021 10:37 a. m.	Archivo de valores...	234 KB
 EFraScrap 1456.csv		05/10/2021 10:37 a. m.	Archivo de valores...	233 KB
 EFraScrap 1457.csv		05/10/2021 10:37 a. m.	Archivo de valores...	232 KB
 EFraScrap 1458.csv		05/10/2021 10:37 a. m.	Archivo de valores...	233 KB
 EFraScrap 1459.csv		05/10/2021 10:37 a. m.	Archivo de valores...	233 KB
 EFraScrap 1460.csv		05/10/2021 10:37 a. m.	Archivo de valores...	232 KB
 EFraScrap 1461.csv		05/10/2021 10:37 a. m.	Archivo de valores...	233 KB
 EFraScrap 1462.csv		05/10/2021 10:37 a. m.	Archivo de valores...	232 KB
 EFraScrap 1463.csv		05/10/2021 10:37 a. m.	Archivo de valores...	232 KB
 EFraScrap 1464.csv		05/10/2021 10:37 a. m.	Archivo de valores...	233 KB
 EFraScrap 1465.csv		05/10/2021 10:37 a. m.	Archivo de valores...	233 KB
 EFraScrap 1466.csv		05/10/2021 10:37 a. m.	Archivo de valores...	234 KB
 EFraScrap 1467.csv		05/10/2021 10:37 a. m.	Archivo de valores...	233 KB
 EFraScrap 1468.csv		05/10/2021 10:37 a. m.	Archivo de valores...	232 KB
 EFraScrap 1469.csv		05/10/2021 10:37 a. m.	Archivo de valores...	232 KB
 EFraScrap 1470.csv		05/10/2021 10:37 a. m.	Archivo de valores...	232 KB
 EFraScrap 1471.csv		05/10/2021 10:38 a. m.	Archivo de valores...	232 KB
 EFraScrap 1472.csv		05/10/2021 10:38 a. m.	Archivo de valores...	233 KB
 EFraScrap 1473.csv		05/10/2021 10:38 a. m.	Archivo de valores...	232 KB
 EFraScrap 1474.csv		05/10/2021 10:38 a. m.	Archivo de valores...	231 KB
 EFraScrap 1475.csv		05/10/2021 10:38 a. m.	Archivo de valores...	231 KB
 EFraScrap 1476.csv		05/10/2021 10:31 a. m.	Archivo de valores...	231 KB
 EFraScrap 1477.csv		05/10/2021 10:31 a. m.	Archivo de valores...	230 KB
 EFraScrap 1478.csv		05/10/2021 10:31 a. m.	Archivo de valores...	230 KB
 EFraScrap 1479.csv		05/10/2021 10:31 a. m.	Archivo de valores...	230 KB
 EFraScrap 1480.csv		05/10/2021 10:31 a. m.	Archivo de valores...	231 KB
 EFraScrap 1481.csv		05/10/2021 10:31 a. m.	Archivo de valores...	230 KB
 EFraScrap 1482.csv		05/10/2021 10:31 a. m.	Archivo de valores...	231 KB
 EFraScrap 1483.csv		05/10/2021 10:31 a. m.	Archivo de valores...	231 KB
 EFraScrap 1484.csv		05/10/2021 10:31 a. m.	Archivo de valores...	230 KB
 EFraScrap 1485.csv		05/10/2021 10:31 a. m.	Archivo de valores...	230 KB

# Metodología

La distribución de carga de trabajo entre el equipo fue la siguiente y fue decidida así desde el inicio de la construcción del presente proyecto:

Crear la cuenta dev	Fase 1: Recopilación de datos de redes sociales	Fer Joel
Hacer el script de Python con Tweepy	Fase 1: Recopilación de datos de redes sociales	Meli Joel
Realizar la recopilación de tweets	Fase 1: Recopilación de datos de redes sociales	Fer Joel
Procesar los CSV's para recopilar todos los tweets y posteriormente distribuirlos	Fase 2: Preprocesamiento de textos	Fer Óscar
Realizar el procedimiento de lemantización	Fase 3: Extracción de características	Efra Meli
Realizar la clasificación de análisis de sentimientos	Fase 4: Clasificación de análisis de sentimientos	Dari Efra
Preparar la arquitectura cliente - servidor	Fase 5: Generación de modelo de Machine Learning en paralelo	Efra Óscar Dari
Preparar un modelo de computación paralela	Fase 5: Generación de modelo de Machine Learning en paralelo	Meli Efra Óscar
Realizar el algoritmo genético para la planificación de carga	Fase 5: Generación de modelo de Machine Learning en paralelo	Dari Fer Joel
Realizar la evaluación de tweets y el procesamiento paralelo	Fase 5: Generación de modelo de Machine Learning en paralelo	Meli Joel Óscar Dari Fer Efra
Documentar el proyecto	Fase Final: Documentar y formalizar el proyecto para su entrega	Dari Meli Óscar

Con una distribución equivalente para todos los integrantes.

# Computación Paralela

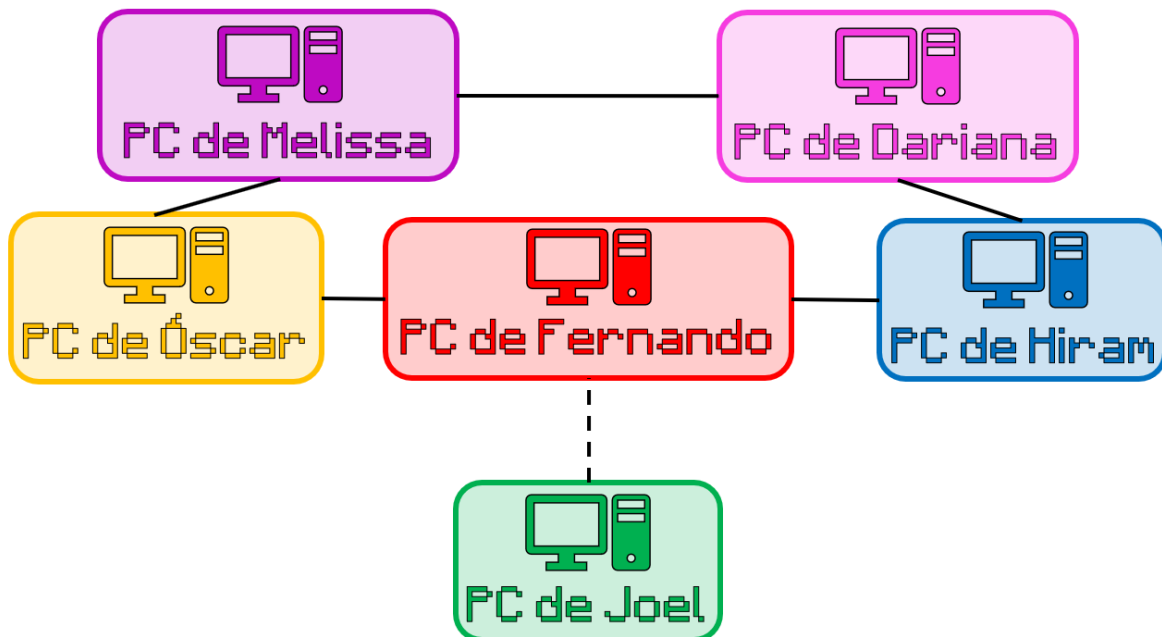
## Explicación de la propuesta Cliente – Servidor del Equipo

La propuesta de arquitectura del equipo se sustenta en dos tipos de archivos base en código de Python: un archivo a nivel de servidor y otro a nivel de cliente.

A partir de ahora y para el resto del documento se denotarán las computadoras como:

- PC de Melissa.
- PC de Joel
- PC de Óscar.
- PC de Dariana.
- PC de Fernando.
- PC de Hiram

La arquitectura propuesta es la siguiente:



La idea de la arquitectura será principalmente un cliente con información a cinco servidores a modo de topología tipo anillo. Para efectos de optimizar más los tiempos y tener un sexto servidor, el servidor también actuará como otro cliente más. En el sentido estricto de que para este trabajo, el servidor cumplirá la función de procesar los tweets y realizar el trabajo que regresará al cliente; para efectos de la actual práctica, se quiere decir, que el cliente también intentará ser un servidor que esté trabajando junto con los demás equipos, pues en términos rígidos del presente proyecto, el servidor tiene como única función realizar la distribución de carga a los demás equipos, es por tanto, que la PC de Joel realizará la tarea de cliente junto con el procesamiento de información, en la que estarán los equipos de los seis integrantes trabajando. En este sentido, los archivos básicos serán dos: el archivo del servidor y el archivo del cliente.

Por la estandarización de nombres pasada, los equipos cumplen las siguientes funciones:

PC de Melissa		Servidor
PC de Joel	Cliente	Servidor
PC de Óscar		Servidor
PC de Dariana		Servidor
PC de Fernando		Servidor
PC de Hiram		Servidor

Para ello, se desarrollaron dos archivos: el archivo en ejecución del servidor y el archivo del cliente.



## El Servidor

El servidor tiene dos fases de código (véase anexo 3), donde la primera esperará la recepción de los otros cinco equipos:

```
### Conexión
import socket

s = socket.socket()
host = socket.gethostname()
port = 8080
s.bind((host,port))
s.listen(1)
print(host)
print("Esperando conexiones...")
conn, addr = s.accept()
print(addr + " se ha conectado")
```

Por este medio, los usuarios se conectarán al servidor, quien esperará su registro en la conexión y posteriormente será notificado el servidor.

Una vez que estén conectados los equipos, el servidor pasará a transmitir archivos de control a los usuarios. Más adelante se explicarán estos archivos de control.

```
### Transfiriendo archivos
filename = "Control.txt"
file = open(filename, 'rb')
file_data = file.read(1024)
conn.send(file_data)
print("Archivo enviado")
```

## El Cliente

Por otra parte, el archivo del cliente (véase anexo 4), será quien espere información del servidor, que al ser proporcionada, la ingresará en el programa.



Así se podrá conectar en el servidor una vez que el servidor comience a ejecutarse:

```
### Conexión
import socket

s = socket.socket()
print("Ingresa la dirección del host")
host = input()
port = 8080
s.connect((host,port))
print("Conectado")
```

Después, el archivo contiene otro módulo de transferencia de datos. Es por este módulo por el que se enviarán datos del servidor al cliente y viceversa.

```
### Transfiriendo archivos
filename = "Control.txt"
file = open(filename, 'wb')
file_data = s.recv(1024)
file.write(file_data)
file.close()
print("Recibido")
```



Los archivos principales por enviar serán los tweets sin analizar del servidor al cliente y éstos mismos analizados del cliente al servidor, sin embargo, otros archivos de control serán enviados entre los equipos.

## ¿Cómo funciona el planificador de carga?


El planificador de carga en sí será todo documento compartido por medio de archivos de formato .txt entre el servidor y el cliente.

Una prueba de trabajo de ellos es el archivo prueba.txt ubicado en el equipo del servidor junto a su respectivo directorio. En este caso de entrega, el archivo no contiene ninguna información relevante de la planificación de tareas, sin



embargo, pueden mostrarse los dos directorios de servidor y cliente respectivamente a continuación:

 prueba.txt	06/10/2021 06:38 p. m.	Documento de te...	1 KB
 server.py	06/10/2021 11:30 p. m.	Archivo PY	1 KB

 clientJoul.py	06/10/2021 11:30 p. m.	Archivo PY	1 KB
---	------------------------	------------	------

Y al ejecutarse ambos programas y estar ambos equipos en contacto, el cliente obtiene este archivo de texto:

 clientJoul.py	06/10/2021 11:30 p. m.	Archivo PY	1 KB
 prueba.txt	06/10/2021 06:38 p. m.	Documento de te...	1 KB

Sin embargo, el verdadero orden de documentos en cliente y servidor será regularmente el siguiente:

- El servidor:
  - El archivo Python del servidor `server.py`.
  - Un archivo de control de planificación de carga y administración `Control-Servidor.txt`.
- El cliente:
  - El archivo Python del servidor `client.py`.
  - Un archivo de recepción de control de planificación de carga y administración `Recepcion-Nombre.txt`. El nombre del archivo variará según el integrante del equipo. Ejemplos: `Recepcion-Melissa.txt`, `Recepcion-Hiram.txt`.
  - Un archivo de emisión de control de planificación de carga y administración `Emission-Nombre.txt`. El nombre del archivo variará según el integrante del equipo. Ejemplos: `Emission-Dariana.txt`, `Emission-Oscar.txt`.

La función de los archivos de texto plano serán encargarse del plan de carga. Al realizarse, los números comenzarán a cambiar en los archivos. Primeramente el

archivo de control del servidor (véase anexo 5) tendrá la información general de los clientes.

El estado del servidor y de cada cliente. 0 significa que está apagado el servidor o los clientes no están en línea y 1 significa que el servidor está activo o los clientes están conectados a la red.

```
Estado: 0
Cliente Melissa: 0
Cliente Joel: 0
Cliente Oscar: 0
Cliente Dariana: 0
Cliente Fernando: 0
Cliente Hiram: 0
```

Por ejemplo, cuando el servidor esté activo, el primer renglón estará activo y los integrantes conectados en ese momento estarán señalados en el archivo también:

```
Estado: 1
Cliente Melissa: 1
Cliente Joel: 0
Cliente Oscar: 0
Cliente Dariana: 0
Cliente Fernando: 1
Cliente Hiram: 1
```

Posteriormente se encuentra un apartado para cada integrante en el que se reporta el progreso personal de cada integrante (véase anexo 6) y la asignación de cada uno a nivel cíclico dinámico como se muestra a continuación:

MELISSA

```
Tweets enviados a procesar: 0
Tweets ya procesados: 0
Tiempo: 0
```

Este formato se repite para los seis participantes del proyecto.

En un ejemplo por dar explicación a cada variable tomada en cuenta para cada usuario, se propone el siguiente ejemplo a explicar:

MELISSA

Tweets enviados a procesar: 400

Tweets ya procesados: 1000

Tiempo: 1.2302

La imagen anterior estaría proporcionando la información del PC de Melissa donde se habría indicado por el algoritmo genético habría calculado que lo óptimo es enviarle en el actual periodo de tiempo 400 tweets a procesar, mientras que ya tendría 1,000 tweets reportados al servidor como procesados en un tiempo de 1.2302 segundos.

Por parte del cliente, estos datos serían enviados al archivo de recepción del cliente (véase anexo 6) en el que se le indicaría primeramente si su conexión ya se realizó y posteriormente los tweets asignados para que trabaje su equipo, junto con un archivo CSV con los tweets indicados.

MELISSA

Conexion: 0

Tweets recibidos para procesar: 0

Cuando el cliente termina de trabajar en su procesamiento, escribirá sobre el archivo de emisión que será enviado al servidor con los datos que actualizará sobre el archivo de control, proporcionando los siguientes datos:

MELISSA

Tweets procesados: 0

Tiempo: 0

De modo que la información del plan de carga se mantendría en flujo en sintonía con los archivos CSV enviados a través de los usuarios de esta red y, donde si todo funciona bien, el planificador de carga mantendría una gestión de cada usuario correcta.

## Distribución de Carga Usando un Algoritmo Genético

Se basó mucho en el artículo proporcionado en clases *“Observations on Using Genetic Algorithms for Dynamic Load-Balancing”* sin embargo, propone una arquitectura muy atractiva para la distribución de carga tanto dinámica como variable, pues en él se expone una idea con tareas de distinto peso.

En el caso propio, el equipo pensó en principio que no existiría esto en la presentación del algoritmo, ya que las tareas a procesar son el análisis de sentimientos de los tweets. Cada tweet deberá pasar por el mismo procedimiento y por ello se consideraba que se analizaría con el mismo peso, ya que se cree que las diferencias de tiempo pueden ser discriminadas. Esto debido a que primero se planteó una asignación de pesos a cada tweet bajo la fórmula:

$$Peso = \ln(Palabras)$$

Pero para la velocidad de los tiempos de ejecución, el comportamiento logarítmico era despreciable, ya que la cantidad máxima de palabras era de 200 en cada tweet, lo que hizo considerar más en despreciar estas diferencias.

Entonces frente a un algoritmo de balanceo de carga de tareas donde cada tarea es un tweet y por lo tanto, cada tarea pesa lo mismo, la pregunta del equipo recaía en cómo realizar este procedimiento interesante, ya que se designó como valor 1 a cada tweet pero sin dar aún una propuesta atractiva en el proceso, lo más sencillo que podía hacerse, incluso sin un algoritmo genético, era repartir todas las tareas entre seis y que el procedimiento siga su curso.

Esta propuesta, aunque cierta, no realizaba ninguna tarea de evaluación heurística, por lo que el equipo discutió que este proyecto consiste en la distribución de carga de modo que un equipo no esté trabajando más de la cuenta mientras que otro equipo carece de trabajo. Para ello se propuso realizar el balance de carga con todas las tareas teniendo el mismo peso pero variando la potencia del procesador

de cada computadora, por lo que primeramente se midió el poder de procesamiento de cada equipo.

Primeramente se diseñó un problema computacionalmente fácil pero que requiriera de mucho tiempo para completarse (véase el anexo 7) y se reportaron los tiempos de ejecución de cada equipo con una serie de ejecuciones.

Después de realizar variadas ejecuciones del problema computacional anteriormente mencionado, se realizó una media de tiempo por PC. Los resultados fueron los siguientes (véase anexo 8 para evidencias de los resultados):

PC de Melissa	9.311
PC de Joel	14.8685
PC de Óscar	7.458
PC de Dariana	90.8248
PC de Fernando	16.1
PC de Hiram	23.006

Con estos datos conocidos, la idea era asignar la cantidad adecuada de tweets para que todos los procesadores, con su debida potencia, procesaran los tweets que les permitieran sus capacidades.

### Definición del algoritmo genético

El algoritmo genético propuesto está planteado bajo los términos siguientes:

- Sea  $q_i$  (*Quantity of Individuals*) una variable entera positiva que representará la cantidad de cromosomas en la población del algoritmo genético.
- Sea *sample* una matriz computacional de tres columnas y  $q_i$  filas:
- Sea *qelitism* (*Quantity of Elitism*) una variable entera positiva que representará la cantidad de individuos que se heredarán por elitismo.

- Sea  $pc$  (*Probability of Crossing*) una variable entera dentro del intervalo  $[0,100]$  que representará la probabilidad de que se active el cruzamiento.
- Sea  $pm$  (*Probability of Mutation*) una variable entera dentro del intervalo  $[0,100]$  que representará la probabilidad de que se active la mutación.
- Sea  $qg$  (*Quantity of Generations*) una variable entera positiva que representará la cantidad de generaciones que se le indicarán al algoritmo genético para repetir el procedimiento.

sample tendrá la siguiente forma:

$$\begin{bmatrix} i_1 & f(i_1) & F(i_1) \\ i_2 & f(i_2) & F(i_2) \\ i_3 & f(i_3) & F(i_3) \\ \vdots & \vdots & \vdots \\ i_{qi} & f(i_{qi}) & F(i_{qi}) \end{bmatrix}$$

Donde  $i_n$  es el n-ésimo cromosoma de la población;  $f(i_n)$  es el resultado de evaluar el n-ésimo cromosoma en la función objetivo – que más adelante se tratará y desarrollará – y  $F(i_n)$  es la función acumulada objetivo desde el primer cromosoma hasta el n-ésimo cromosoma, es decir  $F(i_n)$  está definida como:

$$F(i_n) = \sum_{k=1}^n f(i_k)$$

También cabe aclarar que un cromosoma general tiene una composición compleja, pues todos los cromosomas están definidos como una tupla de seis números de la siguiente forma:

$$i_n = (iM_n \quad iJ_n \quad iO_n \quad iD_n \quad iF_n \quad iH_n)$$

Donde:

- $iM_n$  será la asignación de tweets para Melissa en la n-ésima solución.
- $iJ_n$  será la asignación de tweets para Joel en la n-ésima solución.

- $iO_n$  será la asignación de tweets para Óscar en la n-ésima solución.
- $iD_n$  será la asignación de tweets para Dariana en la n-ésima solución.
- $iF_n$  será la asignación de tweets para Fernando en la n-ésima solución.
- $iH_n$  será la asignación de tweets para Hiram en la n-ésima solución.

La representación anterior está sujeta a la siguiente condición:

$$iM_n + iJ_n + iO_n + iD_n + iF_n + iH_n = total\_tweets$$

Donde `total_tweets` es una variable computacional entera positiva que tendrá almacenado el valor de tweets total a procesar.

Y debido a que de ambas maneras se puede representar un cromosoma, la siguiente igualdad se satisface.

$$sample = \begin{bmatrix} i_1 & f(i_1) & F(i_1) \\ i_2 & f(i_2) & F(i_2) \\ i_3 & f(i_3) & F(i_3) \\ \vdots & \vdots & \vdots \\ i_{qi} & f(i_{qi}) & F(i_{qi}) \end{bmatrix} = \begin{bmatrix} (iM_1 & iJ_1 & iO_1 & iD_1 & iF_1 & iH_1) & f(i_1) & F(i_1) \\ (iM_2 & iJ_2 & iO_2 & iD_2 & iF_2 & iH_2) & f(i_2) & F(i_2) \\ (iM_3 & iJ_3 & iO_3 & iD_3 & iF_3 & iH_3) & f(i_3) & F(i_3) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (iM_{qi} & iJ_{qi} & iO_{qi} & iD_{qi} & iF_{qi} & iH_{qi}) & f(i_{qi}) & F(i_{qi}) \end{bmatrix}$$

Es decir, es igual la representación de los cromosomas aunque se hablará con mayor profundidad usando la segunda representación.

## El procedimiento del algoritmo genético

El procedimiento del algoritmo genético está definido de la siguiente manera:

### Generación de la población inicial

Sea `remaining_tweets` una variable con características similares a `total_tweets`.

El procedimiento inicia estableciendo el valor de `total_tweets` en `remaining_tweets`.



Posteriormente se generará un número aleatorio entre 0 y `remaining_tweets` que se le asignará al primer elemento de la tupla del cromosoma. Asimismo se restará este valor a `remaining_tweets`, valor que se asignará a sí mismo.

Con el valor actualizado, se generará nuevamente otro número aleatorio entre 0 y el nuevo valor de `remaining_tweets` y se le asignará al segundo elemento de la tupla del cromosoma. Nuevamente este valor se restará este valor a `remaining_tweets`, valor que se asignará a sí mismo.

Este procedimiento se repite para los seis elementos de la tupla.

En caso de que `remaining_tweets` no sea cero para cuando se haya terminado de asignar valor a cada uno de los seis elementos, entonces se elegirá un elemento al azar para asignar el resto del valor de `remaining_tweets` a este elemento de la tupla.

Este procedimiento se repite para cada cromosoma dentro de la población.

## Evaluación y Función Objetivo

Denotaremos para este punto los tiempos indicados por cada integrante de equipo como `timeM`, `timeJ`, `timeO`, `timeD`, `timeF` y `timeH` respectivamente.

Para plantear la función objetivo primeramente se pensó que la cantidad de tweets dividido entre el tiempo registrado de una persona deberá ser muy similar a la cantidad de tweets dividido entre el tiempo registrado de otra persona, por ejemplo:

$$\frac{iM_k}{timeM} = \frac{iJ_k}{timeJ} \quad \forall k \in \{1,2,3, \dots, qi\}$$

Esto debe ocurrir para todos los integrantes del equipo:

$$\frac{iM_k}{timeM} = \frac{iJ_k}{timeJ} = \frac{iO_k}{timeO} = \frac{iD_k}{timeD} = \frac{iF_k}{timeF} = \frac{iH_k}{timeH} \quad \forall k \in \{1,2,3, \dots, qi\}$$

La mejor forma para ajustar todos estos valores a uno similar podrá ser la herramienta estadística conocida como la varianza. Para ello primero se calculó la media de todos estos valores:

$$Media = \frac{\left( \frac{iM_k}{timeM} + \frac{iJ_k}{timeJ} + \frac{iO_k}{timeO} + \frac{iD_k}{timeD} + \frac{iF_k}{timeF} + \frac{iH_k}{timeH} \right)}{6}$$

Posteriormente se calculará un valor muy similar a la definición formal de varianza en Estadística Descriptiva:

$$\sigma^2 = \left| \frac{iM_k}{timeM} - Media \right| + \left| \frac{iJ_k}{timeJ} - Media \right| + \left| \frac{iO_k}{timeO} - Media \right| + \left| \frac{iD_k}{timeD} - Media \right| + \left| \frac{iF_k}{timeF} - Media \right| + \left| \frac{iH_k}{timeH} - Media \right|$$

Ahora, el valor más pequeño de  $\sigma^2$  será el que describa mejor a un cromosoma, pero por fines prácticos del grupo, se conoce mejor la forma de evaluarlos cuando una función expresa a un cromosoma como “mejor” cuando ésta le asigna un valor mayor, para ello se invirtió la función:

$$Función\ objetivo = \frac{1}{\sigma^2}$$

Finalmente, se tuvieron pequeños inconvenientes en el desarrollo del código que llevaron a corregir la función debido a sus asíntotas, de modo que la función final será la siguiente:

$$Función\ objetivo = \frac{1}{(\sigma^2 \times 1,000) + 1} \times 100$$

Esta misma función califica a los cromosomas, por lo que con ella, pueden ordenarse los cromosomas y realizar elitismo. En tal caso, el algoritmo tomará los *qelitism* mejores que serán directamente insertados en la nueva población.

## Etapa de Cruzamiento

De la misma forma como la función objetivo es útil para reportar a los mejores cromosomas y aplicar elitismo, ésta también funciona para realizar la selección por ruleta, que es una selección por ruleta tradicional.

Una vez que se seleccionan dos cromosomas, el siguiente operador a ejecutar será el cruzamiento, que será difícil ejecutar debido al modelo planteado.

Sean las tuplas  $(iM_h \ iJ_h \ iO_h \ iD_h \ iF_h \ iH_h)$  y  $(iM_k \ iJ_k \ iO_k \ iD_k \ iF_k \ iH_k)$  las elegidas por ruleta para realizar cruzamiento.

La primera idea era adaptar el cruzamiento tradicional del algoritmo genético simple. Esto implica realizar un punto de corte en ambas tuplas y cambiar los valores. Es decir, un posible resultado de las tuplas podía ser el siguiente:

- $(iM_h \ iJ_h \ iO_k \ iD_k \ iF_k \ iH_k)$
- $(iM_k \ iJ_k \ iO_h \ iD_h \ iF_h \ iH_h)$

El problema es que este cruzamiento no garantiza que la condición de suma (donde la suma de todos los elementos de una tupla resulte en el total de tweets) se satisfaga después de aplicar el operador, por lo que la idea se descartó.

Sin embargo, en lugar de realizar un punto de corte, la idea anterior le proporcionó al equipo una segunda propuesta que podría ser la definitiva, pues en lugar de realizar un punto de corte, quizá lo mejor sería sólo intercambiar un miembro de cada tupla, por ejemplo, el resultado sería el siguiente:

- $(iM_h \ iJ_h \ iO_k \ iD_h \ iF_h \ iH_h)$
- $(iM_k \ iJ_k \ iO_h \ iD_k \ iF_k \ iH_k)$

Nuevamente existe el problema que la condición de suma no se cumplirá después de aplicar el cruzamiento, puesto que, estrictamente deberán ser números distintos para que el cruzamiento tenga sentido, sin embargo, la diferencia de valor entre un elemento y otro causará que en una tupla exista un excedente de tweets a la cantidad total y en el otro elemento habrá una falta de tweets para alcanzar la cantidad total, por lo que la idea sería también descartada.

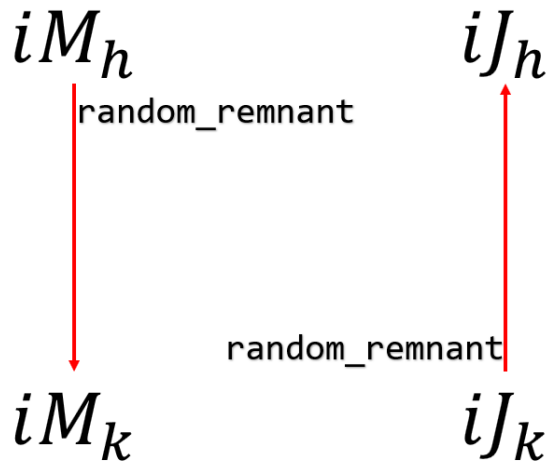
Sin embargo, esta idea anterior llevó a la propuesta final para el cruzamiento, pues esta diferencia podría recuperarse en otro elemento de la tupla, es entonces que se planteó la siguiente propuesta, que fue implementada en la realización del presente documento:

Supóngase que se eligen aleatoriamente los primeros dos elementos de la tupla para intercambiarse:  $iM_h$ ,  $iJ_h$ ,  $iM_k$  e  $iJ_k$ .

Se buscará el valor mínimo de estos valores:

$$top = \min(iM_h, iJ_h, iM_k, iJ_k)$$

Posteriormente se elegirá un número aleatorio entre 0 y  $top$  que se denotará como `random_remnant`. Este valor rotará entre los elementos elegidos:



Es decir, después del cruzamiento, las tuplas tendrán los siguientes valores:

- $(iM_h - random\_remnant \quad iJ_h + random\_remnant \quad iO_h \quad iD_h \quad iF_h \quad iH_h)$
- $(iM_k + random\_remnant \quad iJ_k - random\_remnant \quad iO_k \quad iD_k \quad iF_k \quad iH_k)$

Así, es claro que la cerradura de suma se mantiene para ambos cromosomas.

El modelo de cruzamiento está inspirado en una técnica Simplex de Investigación de Operaciones para balanceo de cargas en el método mencionado anteriormente. Todos los números de las tuplas se combinarán para crear dos hijos.

Este proceso se realiza tres veces (cada vez involucra a dos elementos de una tupla, de modo que hacerlo tres veces involucra a todos los elementos de la tupla) para cada cromosoma siempre que el cruzamiento se active.

### Etapas de Mutación

La mutación realiza un procedimiento similar. En este caso se toma el número menor de la tupla y se generará un número aleatorio entre 0 y el menor de toda la tupla y se rotará este número que se denotará como *add*.

Sea  $(iM_h \quad iJ_h \quad iO_h \quad iD_h \quad iF_h \quad iH_h)$  la tupla que activa el proceso de mutación.

Después de realizar la mutación, la tupla tendrá el siguiente aspecto:

$$(iM_h + add \quad iJ_h - add \quad iO_h + add \quad iD_h - add \quad iF_h + add \quad iH_h - add)$$

Al terminar el procedimiento de mutación, todos los operadores del algoritmo genético habrían sido aplicados, por lo que, si se especificó realizar otra generación, la iteración regresaría a repetir este proceso, de lo contrario, el algoritmo genético habría terminado.

Ahora se llevó lo anteriormente explicado a código Python (véase anexo 9) y lo primero que se realizó fue calibrar el algoritmo genético:

```
## Calibrar variables
total_tweets = 6942021
qe = 6 # Inamovible
qi = 60 # Movible
pc = 75 # Movible
pm = 40 # Movible
qelitism = 5 # Movible
qg = 10
sample = [[], [], []]
power_process = [9.311, 14.8685, 7.458, 90.8248, 16.1, 23.006] # Orden alfabético: Melissa, Joel, Óscar, Dariana,
ig = 0
```

Posteriormente se crea la población inicial y se realizan las evaluaciones con respecto a la función objetivo:

```
## AG: Creación de la población aleatoria
sample[0] = FirstSample(qe,qi,total_tweets)
PrintSample(qi, sample)

## AG: Procedimiento de cada generación
while ig < qg:
    print("Población " + str(ig))
    print("")

    # Evaluamos a todas las muestras
    sample[1] = evaluate(qi,sample,power_process)

    print("Población " + str(ig) + " al ser evaluada")
    PrintSample(qi,sample)
    print("")

    # Evaluamos la suma acumulada
    sample[2] = evaluateA(qi,sample)

    print("Población " + str(ig) + " al ser evaluada con acumulación")
    PrintSample(qi,sample)
    print("")

    # Ordenamos por valor de f(x) es decir, por sample[1]
    sample = Sorting(sample, qi)

    print("Población " + str(ig) + " al ser ordenada")
    PrintSample(qi,sample)
    print("")

    # Reevaluamos la suma acumulada
    sample[2] = evaluateA(qi,sample)

    print("Población " + str(ig) + " al ser reevaluada")
    PrintSample(qi,sample)
    print("")
```

Luego se crea una nueva estructura para la nueva generación y se aprovecha para pasar por elitismo a los mejores individuos:

```
# Creamos una nueva muestra vacía
newSample = newSampleCreator(qi)

print("Nueva muestra " + str(ig + 1) + " vacía")
print(newSample)
print("")

print("Se aplicará elitismo")
x = input()

# Se aplica elitismo
ii = 0
while ii < qelitism:
    newSample[0][ii] = sample[0][ii]
    #newSample[1][ii] = sample[1][ii]
    #newSample[2][ii] = sample[2][ii]
    ii = ii + 1

print("Nueva muestra " + str(ig + 1) + " después de elitismo")
for i in range(qi):
    try:
        print(str(newSample[0][i]) + " = " + str(sum(newSample[0][i])))
    except:
        pass
print("")
print("Se aplicará cruzamiento")
x = input()
```

Después se realiza el cruzamiento primeramente realizando la selección de dos padres:

```
# Se aplicará cruzamiento
while ii < qi:
    # Seleccionamos por ruleta
    # Padre 1
    random_pick = random.uniform(0, sample[2][qi - 1])

    for ic in range(qi):
        if random_pick < sample[2][ic]:
            c1 = sample[0][ic]
            print("Se elige como padre 1 el individuo " + str(ic) + "(random_pick = " + str(random_pick) + ")")
            break

    # Padre 2
    random_pick = random.uniform(0, sample[2][qi - 1])

    for ic in range(qi):
        if random_pick < sample[2][ic]:
            c2 = sample[0][ic]
            print("Se elige como padre 2 el individuo " + str(ic) + "(random_pick = " + str(random_pick) + ")")
            break
```

Lo siguiente es evaluar si los padres elegidos realizarán cruzamiento o no:

```
ii = 0
while ii < qi:
    # Se evalúa si se aplicará mutación
    random_num = random.randint(0,99)
    if random_num < pm:
        print("El cromosoma " + str(ii) + " ha activado mutación (Probabilidad: " + str(random_num) +
        print("")

        processors = [0, 1, 2, 3, 4, 5]
        random.shuffle(processors) # Se mezclarán y se hará una sumas balanceadas cíclica

        print("Tupla mezclada: " + str(processors))
        print("")

        print("Valores antes de la mutación: ")
        print(newSample[0][ii])
        print("")

        min_value = min(newSample[0][ii])
        operation = random.randint(0, 1) # Elige la operación con la que empezará la suma equilibrada
        addition = random.randint(0, min_value)

        print("Tupla " + str(processors))
        print("min_value = " + str(min_value))
        print("addition = " + str(addition))
        print("")
```

Posteriormente se harán lo que en el equipo se le denominó “sumas equilibradas”, que es el concepto de equilibrio por medio de una técnica Simplex:

```
for i in [0, 2, 4]:
    # Sumas equilibradas
    min_value = min(c1[processors[i]], c1[processors[i + 1]], c2[processors[i]], c2[processors[i + 1]])
    operation = random.randint(0, 1) # Elige la operación con la que empezará la suma equilibrada
    addition = random.randint(0, min_value)

    print("Tuplas " + str(processors[i] + 1) + " y " + str(processors[i + 1] + 1))
    print("min_value = " + str(min_value))
    print("addition = " + str(addition))
    print("")

    if operation == 0:
        c2[processors[i + 1]] = c2[processors[i + 1]] - addition
        c1[processors[i + 1]] = c1[processors[i + 1]] + addition
        c1[processors[i]] = c1[processors[i]] - addition
        c2[processors[i]] = c2[processors[i]] + addition
    else:
        c2[processors[i + 1]] = c2[processors[i + 1]] + addition
        c1[processors[i + 1]] = c1[processors[i + 1]] - addition
        c1[processors[i]] = c1[processors[i]] + addition
        c2[processors[i]] = c2[processors[i]] - addition

    print("Valores después del cruzamiento: ")
    print(c1)
    print(c2)
    print("")
```

Si no se cruzan, el procedimiento contrario es heredarlos de igual manera.



Después se realiza la mutación:

```
# Se aplica mutación
ii = 0
while ii < qi:
    # Se evalúa si se aplicará mutación
    random_num = random.randint(0,99)
    if random_num < pm:
        print("El cromosoma " + str(ii) + " ha activado mutación (Probabilidad: " + str(random_num) + " < "
        print("")

        processors = [0, 1, 2, 3, 4, 5]
        random.shuffle(processors) # Se mezclarán y se hará una sumas balanceadas cíclica

        print("Tupla mezclada: " + str(processors))
        print("")

        print("Valores antes de la mutación: ")
        print(newSample[0][ii])
        print("")

        min_value = min(newSample[0][ii])
        operation = random.randint(0, 1) # Elige la operación con la que empezará la suma equilibrada
        addition = random.randint(0, min_value)

        print("Tupla " + str(processors))
        print("min_value = " + str(min_value))
        print("addition = " + str(addition))
        print("")

        if operation == 0:
            newSample[0][ii][processors[0]] = newSample[0][ii][processors[0]] - addition
            newSample[0][ii][processors[1]] = newSample[0][ii][processors[1]] + addition
            newSample[0][ii][processors[2]] = newSample[0][ii][processors[2]] - addition
            newSample[0][ii][processors[3]] = newSample[0][ii][processors[3]] + addition
            newSample[0][ii][processors[4]] = newSample[0][ii][processors[4]] - addition
            newSample[0][ii][processors[5]] = newSample[0][ii][processors[5]] + addition
```

El procedimiento se repetirá según la cantidad de generaciones deseado.

Un ejemplo de ejecución de código es el siguiente. A continuación se muestra la población inicial al ser evaluada y ordenada:

Población 0 al ser reevaluada		
[2969361, 1539170, 1815651, 18523, 412987, 186329] = 6942021	31622037	31622037
[1865628, 1340774, 3321758, 89961, 257220, 66680] = 6942021	31562903	63184940
[2199470, 2560613, 1966710, 93170, 32732, 89326] = 6942021	31546451	94731391
[758906, 3125166, 1812820, 347267, 419893, 477969] = 6942021	30975301	125706692
[668383, 1823891, 2580073, 1260065, 74070, 535539] = 6942021	30969025	156675717
[2547676, 1679639, 1001566, 977186, 302243, 433711] = 6942021	30870443	187546160
[875519, 2158420, 3016210, 586654, 191857, 113361] = 6942021	30774515	218320675
[1845586, 3211955, 444884, 202954, 1088166, 148476] = 6942021	30756221	249076896
[1680857, 1036057, 3977078, 218624, 2748, 26657] = 6942021	30399821	279476717
[2650913, 2797244, 147921, 126160, 717987, 501796] = 6942021	30170303	309647020
[227888, 2168853, 2913690, 981996, 641130, 8464] = 6942021	30168005	339815025
[2948873, 2369384, 816626, 341442, 283551, 182145] = 6942021	30106627	369921652
[3583567, 791277, 1486148, 492479, 398246, 190304] = 6942021	29768075	399689727
[568776, 723038, 3902812, 558178, 421514, 767703] = 6942021	29752731	429442458
[3441356, 688951, 2009969, 652667, 136069, 13009] = 6942021	29732121	459174579
[3793563, 313321, 2046431, 372089, 361754, 54863] = 6942021	29588863	488763442
[2551365, 3115234, 757508, 258417, 197677, 61820] = 6942021	29575993	518339435

[2482156, 3272432, 934092, 193415, 44168, 15758] = 6942021	29530019 547869454
[985930, 972146, 1833844, 2523668, 351808, 274625] = 6942021	29366973 577236427
[2650989, 2815668, 873007, 598450, 2346, 1561] = 6942021	29295811 606532238
[4078958, 1395460, 1084937, 340042, 507, 42117] = 6942021	29082167 635614405
[223062, 664049, 2854577, 1731647, 371117, 1097569] = 6942021	28842531 664456936
[4110636, 864439, 1453124, 461013, 42878, 9931] = 6942021	28776869 693233805
[4405407, 1578050, 554754, 168294, 221645, 13871] = 6942021	28749945 721983750
[3165453, 1296076, 764776, 1425381, 279518, 10817] = 6942021	28738499 750722249
[3678398, 28387, 1106706, 736053, 1260842, 131635] = 6942021	28597099 779319348
[4490396, 1031536, 1239910, 73570, 99624, 6985] = 6942021	28579967 807899315
[2606706, 1447308, 488702, 2090652, 26763, 281890] = 6942021	28525451 836424766
[4574472, 1273115, 925421, 9432, 104218, 55363] = 6942021	28411815 864836581
[880266, 4004944, 904777, 792508, 357802, 1724] = 6942021	28325263 893161844
[4617886, 656977, 1571440, 23067, 52409, 20242] = 6942021	28324987 921486831
[445283, 939017, 4049558, 1292680, 176507, 38976] = 6942021	28106749 949593580
[202769, 3178837, 1264093, 1763174, 124357, 408791] = 6942021	28036145 977629725
[3758510, 2760947, 317397, 22715, 40287, 42165] = 6942021	28027703 1005657428
[4841075, 1390739, 295531, 202586, 140523, 71567] = 6942021	27832845 1033490273
[4935908, 540620, 825644, 165224, 459982, 14643] = 6942021	27688943 1061179216
[3679508, 3014407, 108490, 35119, 12794, 91703] = 6942021	27678787 1088858003
[4331043, 1039278, 725498, 792389, 49251, 4562] = 6942021	27673303 1116531306
[88008, 5093751, 1741794, 1174, 7228, 10066] = 6942021	27373257 1143904563
[5349748, 144004, 1408908, 28851, 670, 9840] = 6942021	26861263 1170765826
[463477, 4867792, 253975, 698397, 257604, 400776] = 6942021	26787789 1197553615
[1322455, 5121094, 563, 480577, 3847, 13485] = 6942021	26716825 1224270440

Observemos que todas las tuplas suman 6,942,021. El número siguiente es su valor como función objetivo y el siguiente se trata del valor acumulado para realizar elitismo, que de hecho, al aplicar elitismo en esta población se obtiene lo siguiente:

```
Nueva muestra 1 vacía
[[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0], [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]]

Se aplicará elitismo

Nueva muestra 1 después de elitismo
[2969361, 1539170, 1815651, 18523, 412987, 186329] = 6942021
[1865628, 1340774, 3321758, 89961, 257220, 66680] = 6942021
[2199470, 2560613, 1966710, 93170, 32732, 89326] = 6942021
[758906, 3125166, 1812820, 347267, 419893, 477969] = 6942021
[668383, 1823891, 2580073, 1260065, 74070, 535539] = 6942021
```

Posteriormente en el cruzamiento, observemos cómo el algoritmo avisa que no se ha activado el cruzamiento:

```
Se elige como padre 1 el individuo 3(random_pick = 110888477.71224487)
Se elige como padre 2 el individuo 17(random_pick = 542848757.1963074)
Los cromosomas 49 y 50 no han activado cruzamiento (Probabilidad: !82 < 75!)
```

Asimismo, se aprecia a continuación la activación del cruzamiento:

```
Se elige como padre 1 el individuo 53(random_pick = 1515027406.0065634)
Se elige como padre 2 el individuo 43(random_pick = 1270469611.342023)
Los cromosomas 53 y 54 han activado cruzamiento (Probabilidad: 54 < 75)
```

Cuando esto sucede, se reparten en pares los valores de las tuplas para mezclarse como se muestra a continuación:

```
Tupla mezclada: [4, 2, 3, 5, 0, 1]

Valores antes del cruzamiento:
[445283, 939017, 4049558, 1292680, 176507, 38976]
[4394701, 1559395, 680532, 186949, 95867, 24577]

Tuplas 5 y 3
min_value = 95867
addition = 10059

Tuplas 4 y 6
min_value = 24577
addition = 13195

Tuplas 1 y 2
min_value = 445283
addition = 344965

Valores después del cruzamiento:
[100318, 1283982, 4039499, 1279485, 186566, 52171]
[4739666, 1214430, 690591, 200144, 85808, 11382]
```

Generando una población de individuos ya cruzados:

```
[2910289, 2355836, 817043, 345955, 322135, 190763] = 6942021
[1321644, 5129616, 50, 481090, 4658, 4963] = 6942021
[668843, 1826377, 2574342, 1270469, 71584, 530406] = 6942021
[668843, 1826377, 2574342, 1270469, 71584, 530406] = 6942021
[907762, 3985833, 877281, 793891, 376913, 341] = 6942021
[2446418, 3264221, 961149, 219354, 25496, 25383] = 6942021
[668843, 1826377, 2574342, 1270469, 71584, 530406] = 6942021
[6311967, 610862, 2571, 2131, 1055, 13435] = 6942021
[2456600, 3354140, 760853, 19511, 292442, 58475] = 6942021
[3536121, 450045, 2006624, 891573, 41304, 16354] = 6942021
[5559159, 725333, 625816, 3467, 11910, 16336] = 6942021
[5422971, 70781, 1408387, 28441, 1191, 10250] = 6942021
[6295358, 66474, 359857, 192163, 27896, 273] = 6942021
[3733768, 373116, 2115273, 303247, 361854, 54763] = 6942021
[2910289, 2355836, 817043, 345955, 322135, 190763] = 6942021
[3839643, 2758954, 275755, 14468, 44311, 8890] = 6942021
[3532215, 2699599, 297181, 225772, 138873, 48381] = 6942021
[4987467, 1698043, 97250, 19437, 15345, 124479] = 6942021
[5801053, 302214, 475984, 353544, 9219, 7] = 6942021
[2456600, 3354140, 760853, 19511, 292442, 58475] = 6942021
[3472949, 1287102, 457280, 1414743, 288492, 21455] = 6942021
[5280471, 121154, 731806, 368189, 128, 440273] = 6942021
[4739666, 1214430, 690591, 200144, 85808, 11382] = 6942021
[3689104, 47042, 980928, 717398, 1386620, 120929] = 6942021
[4987467, 1698043, 97250, 19437, 15345, 124479] = 6942021
[5280471, 121154, 731806, 368189, 128, 440273] = 6942021
[5752105, 1131588, 35261, 17545, 910, 4612] = 6942021
[2446418, 3264221, 961149, 219354, 25496, 25383] = 6942021
[6295358, 66474, 359857, 192163, 27896, 273] = 6942021
[784261, 2289257, 2885373, 631150, 213997, 137983] = 6942021
[6533076, 329256, 7074, 36297, 36281, 37] = 6942021
[1844933, 3214720, 442119, 187687, 1103433, 149129] = 6942021
[3839643, 2758954, 275755, 14468, 44311, 8890] = 6942021
[5801053, 302214, 475984, 353544, 9219, 7] = 6942021
[6295358, 66474, 359857, 192163, 27896, 273] = 6942021
[3733768, 373116, 2115273, 303247, 361854, 54763] = 6942021
[755218, 659469, 3578184, 581861, 408772, 958517] = 6942021
[6047208, 22261, 844701, 24559, 3287, 5] = 6942021
[758906, 3125166, 1812820, 347267, 419893, 477969] = 6942021
[2446418, 3264221, 961149, 219354, 25496, 25383] = 6942021
[5926199, 527918, 274824, 53085, 158711, 1284] = 6942021
[755218, 659469, 3578184, 581861, 408772, 958517] = 6942021
```

Para la mutación pasa algo similar, pues se puede apreciar el caso de cuando no se activa la mutación como cuando sí se activa a continuación:

Cuando se activa la mutación, el cromosoma se desordena y un valor será intercambiado entre todos los números de la tupla:

```
Tupla mezclada: [1, 3, 5, 0, 4, 2]

Valores antes de la mutación:
[947291, 467396, 3386111, 389788, 600845, 1150590]

Tupla [1, 3, 5, 0, 4, 2]
min_value = 389788
addition = 33411

Valores después de la mutación:
[980702, 433985, 3419522, 423199, 567434, 1117179]
```

Una vez que terminan estos procedimientos de ejecutarse, la nueva generación cambia de estructura para repetir y volver a heredar las veces que se hayan indicado.

El programa es muy inestable en sus ejecuciones, sin embargo en los mejores retornos se realizaron promedios y cálculos de estos resultados, obteniendo resultados que más adelante se expondrán.



## Procedimiento de Lemantización

En contraste con la derivación, la lematización es mucho más poderosa, va más allá de la reducción de palabras y considera el vocabulario completo de un idioma para aplicar un análisis morfológico a las palabras, con el objetivo de eliminar solo las terminaciones flexivas y devolver la forma base o de diccionario de una palabra, que se conoce como lema.

En sí, de la palabra o frase original sacamos la función de la palabra raíz (lema), para agilizar el proceso se recomienda primero hacer una extracción de palabras centrales y reunir las, después hacer una conversión del tiempo y pasar todo a presente, y para terminar trasladamos de plural a singular todo.

Para realizar la lematización utilizamos el enfoque “WordNet(con etiqueta POS)”, Wordnet es una base de datos léxica disponible públicamente en más de 200 idiomas que proporciona relaciones semánticas entre sus palabras.

Está presente en la biblioteca nltk en Python, Wordnet vincula las palabras en relaciones semánticas por ejemplo sinónimos los cuales agrupa en forma de synsets (un grupo de elementos de datos que son semánticamente equivalentes).

Para utilizarlo se debe de descargar el paquete de nltk, posteriormente hay que descargar Wordnet desde nltk

```
import nltk
from nltk.stem import WordNetLemmatizer
nltk.download('wordnet')
nltk.download('averaged_perceptron_tagger')
nltk.download('punkt')
from nltk.corpus import wordnet
```

Continuamos con el programa inicializando el lematizador, importando pandas y empezando a leer nuestra base de datos de tweets “Team10 Scrapping (Primera Limpieza).csv”.

```
8 lemmatizer = WordNetLemmatizer()
9 import pandas as pd
10 datos=pd.read_csv('Team10 Scrapping (Primera Limpieza).csv')
```

Para evitar que maneje todo como sustantivos utilizamos la etiquetas POS(Part of Speech), la cual sirve para definir el type de una palabra para ver si es un adjetivo, verbo, sustantivo o adverbio.

```
12 def pos_tagger(nltk_tag):
13     if nltk_tag.startswith('J'):
14         return wordnet.ADJ
15     elif nltk_tag.startswith('V'):
16         return wordnet.VERB
17     elif nltk_tag.startswith('N'):
18         return wordnet.NOUN
19     elif nltk_tag.startswith('R'):
20         return wordnet.ADV
21     else:
22         return None
```

Ahora seleccionamos y recorremos la columna que contiene los tweets, para extraerlos y guardarlos en “sentence” para posteriormente aplicar lo que es la lematización.

```
24 archivo="Lemantizacion.csv"
25 csv=open(archivo,"w")
26 for a in range(len(datos.iloc[:,7])):
27
28     sentence = datos.iloc[a,7]
29
30     pos_tagged = nltk.pos_tag(nltk.word_tokenize(sentence))
31
32     print(pos_tagged)
```

En esta parte es donde se hace la clasificación y se exporta en el nuevo archivo de Excel.

```
35     wordnet_tagged = list(map(lambda x:(x[0], pos_tagger(x[1])),
                                pos_tagged))
36     print(wordnet_tagged)
37
38     lemmatized_sentence = []
39
40     for word, tag in wordnet_tagged:
41         if tag is None:
42
43             lemmatized_sentence.append(word)
44         else:
45
46             lemmatized_sentence.append(lemmatizer.lemmatize(word, tag))
47     lemmatized_sentence = " ".join(lemmatized_sentence)
48     lemmatized_sentence = lemmatized_sentence + "\n"
49     try:
50         csv.write(lemmatized_sentence)
51     except ValueError:
52         print("")
```



### Ejemplo de ejecución:

```
[nltk_data] Downloading package wordnet to
[nltk_data]   C:\Users\zente\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]   C:\Users\zente\AppData\Roaming\nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]   date!
[nltk_data] Downloading package punkt to
[nltk_data]   C:\Users\zente\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
Traceback (most recent call last):
```

Primero hace las descargas.

```
(('was', 'VBD'), ('Death', 'NNP'), ('And', 'CC'), ('the', 'DT'), ('word', 'NN'), ('was', 'VBD'), ('nigger', 'JJR'), ('And', 'CC'), ('the', 'DT'), ('word', 'NN'), ('was', 'VBD'), ('death', 'NN'), ('to', 'TO'), ('all', 'DT'), ('niggers', 'NNS'), ('And', 'CC'), ('the...', 'NN'))
[(('RT', 'n'), ('@', 'n'), ('jatella', 'n'), (':', None), ('"', 'n'), ('In', None), ('the', None), ('beginning', 'n'), ('was', 'v'), ('the', None), ('word', 'n'), ('And', None), ('the', None), ('word', 'n'), ('was', 'v'), ('Death', 'n'), ('And', None), ('the', None), ('word', 'n'), ('was', 'v'), ('nigger', 'a'), ('And', None), ('the', None), ('word', 'n'), ('was', 'v'), ('death', 'n'), ('to', None), ('all', None), ('niggers', 'n'), ('And', None), ('the...', 'n'))]
[(('@', 'RB'), ('3privz', 'CD'), ('WTF', 'NNP'), ('IM', 'NNP'), ('JUST', 'NNP'), ('A', 'NNP'), ('NIGGER', 'NNP'), ('THO', 'NNP'))]
[(('@', 'r'), ('3privz', None), ('WTF', 'n'), ('IM', 'n'), ('JUST', 'n'), ('A', 'n'), ('NIGGER', 'n'), ('THO', 'n'))]
```

Después descompone todo por palabras y las analiza.

@ Mattimatikus @ Urbanistbhd @ DracheOhneLP ich nenne andere nigger und werde nicht gesperrrt
I be go to sue Delta chinese chink for not honor NSA who say i can fly . Life or death ! God famed Democrat leave Obama nigger human trafficking ! Law suit ! Europe already
Gon na take everything from Faye dad not to call Teddy a nigger
@ Big0047 Wakikaa kafanaa naaa ma nigger
@ matthew47138586 @ Oscaranking2 @ _littlehuman_ This nigger about to be a father http ://t.co/ydDVoNfwgA
@ akintonm! especially si more than spelling .
@ CashWunner nigger mode well
If he can ' t help market us then fuck him market himself Typical nigger

Y los resultados son exportados a un nuevo archivo de Excel llamado "Lemantización.csv"

## Resultados

El algoritmo genético estaba diseñado para poder configurar la cantidad de tweets sin realizar grandes cambios al código tan sólo con cambiar el valor de la variable `total_tweets`.

Debido a que el algoritmo genético presenta inestabilidades muy corregibles en el cruzamiento y mutación, se ejecutaron múltiples veces y los siguientes fueron los resultados de promedio que arrojó el algoritmo genético:

Tweets totales	Asignados a Melissa	Asignados a Joel	Asignados a Óscar	Asignados a Dariana	Asignados a Fer	Asignados a Hiram
10,000	2,315	1,791	3,209	203	1,234	1,248
20,000	4,992	3,223	6,248	573	2,851	2,113
30,000	7,479	4,841	9,477	758	4,320	3,125
40,000	9,946	6,479	12,638	1,008	5,742	4,187
50,000	12,689	7,843	15,751	1,305	7,364	5,048
60,000	15,363	9,275	19,037	1,431	8,809	6,085
70,000	17,678	11,067	22,198	1,681	10,116	7,260
80,000	20,211	12,640	25,030	2,261	11,840	8,018
90,000	22,770	14,187	28,425	2,277	13,227	9,114
100,000	25,260	15,804	31,722	2,392	14,667	10,155
200,000	50,639	31,489	63,143	5,084	29,186	20,459
300,000	75,594	47,598	94,727	7,613	43,845	30,623
400,000	100,890	63,365	126,235	10,219	58,713	40,578
500,000	126,154	79,165	157,647	12,920	72,799	51,315
600,000	151,388	94,995	189,144	15,535	87,651	61,287
700,000	176,708	110,738	220,767	18,028	102,024	71,735
800,000	201,880	126,630	252,286	20,622	117,026	81,556
900,000	227,448	142,126	283,753	23,267	131,309	92,097
1,000,000	252,667	157,971	315,164	25,970	146,190	102,038
2,000,000	505,063	316,212	630,578	51,690	292,122	204,335
3,000,000	757,480	474,432	945,761	77,645	438,003	306,679

4,000,000	1,009,975	632,475	1,260,855	103,782	584,042	408,871
5,000,000	1,262,458	790,730	1,576,360	129,311	730,179	510,962
6,000,000	1,515,276	948,548	1,891,532	155,274	876,172	613,198
6,942,021	1,752,929	1,097,724	2,188,458	179,704	1,013,759	709,447

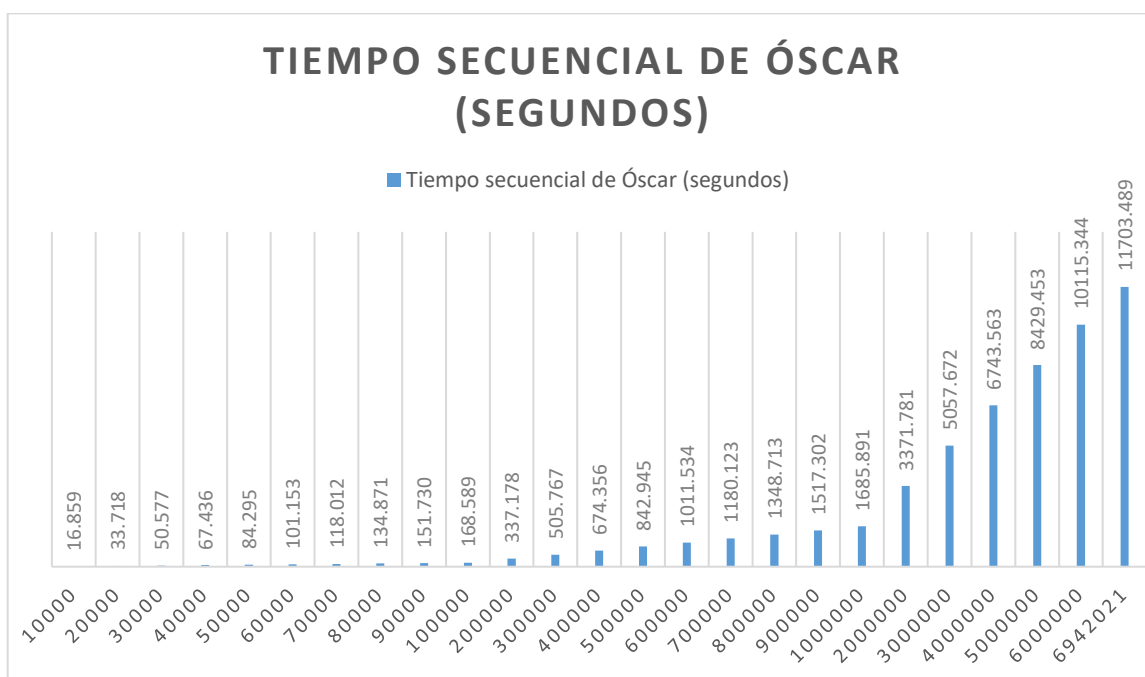
## Resultados de tiempo secuencial

Para el reporte de tiempo secuencial se usará el mejor de los tiempos secuenciales, que en este caso será el tiempo establecido por Óscar (véase anexo 10 para un mayor reporte de tiempos por integrante) que es el siguiente:

Tweets	Tiempo de Óscar (segundos)
10,000	16.859
20,000	33.718
30,000	50.577
40,000	67.436
50,000	84.295
60,000	101.153
70,000	118.012
80,000	134.871
90,000	151.730
100,000	168.589
200,000	337.178
300,000	505.767
400,000	674.356
500,000	842.945
600,000	1011.534
700,000	1180.123
800,000	1348.713
900,000	1517.302

1,000,000	1685.891
2,000,000	3371.781
3,000,000	5057.672
4,000,000	6743.563
5,000,000	8429.453
6,000,000	10115.344
6,942,021	11703.489

Obsérvese la siguiente gráfica

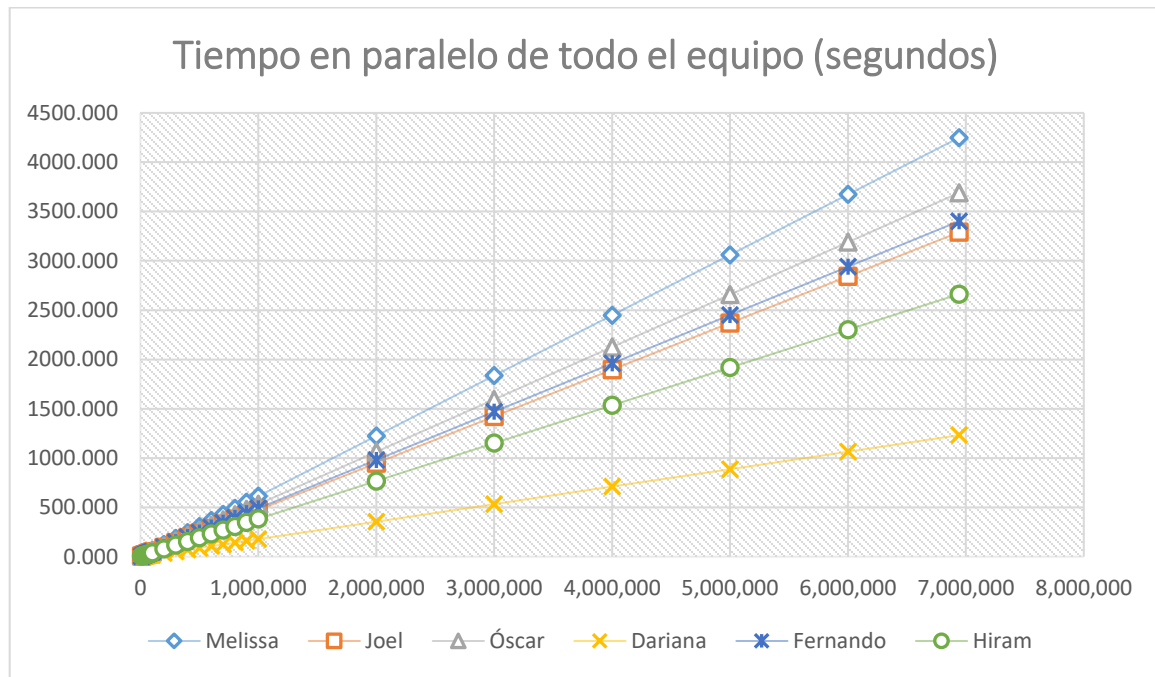


En la gráfica se muestra el crecimiento del tiempo según la cantidad de tweets asignada a nivel secuencial para Óscar. Entre otros datos que pueden observarse en el anexo 10, el PC de Óscar terminaría la tarea de procesar todos los tweets extraídos en tres horas y 15 minutos aproximadamente, si se trata de realizar un procedimiento secuencial, sin embargo, otros equipos como el PC de Melissa tomarían 4 horas y media en realizar toda la tarea secuencial; el PC de Hiram tardaría 7 horas y 10 minutos por sí solo. El peor de los casos sería en el

PC de Dariana, que tardaría 13 horas y 10 minutos terminar de procesar los más de seis millones de tweets que se almacenaron.

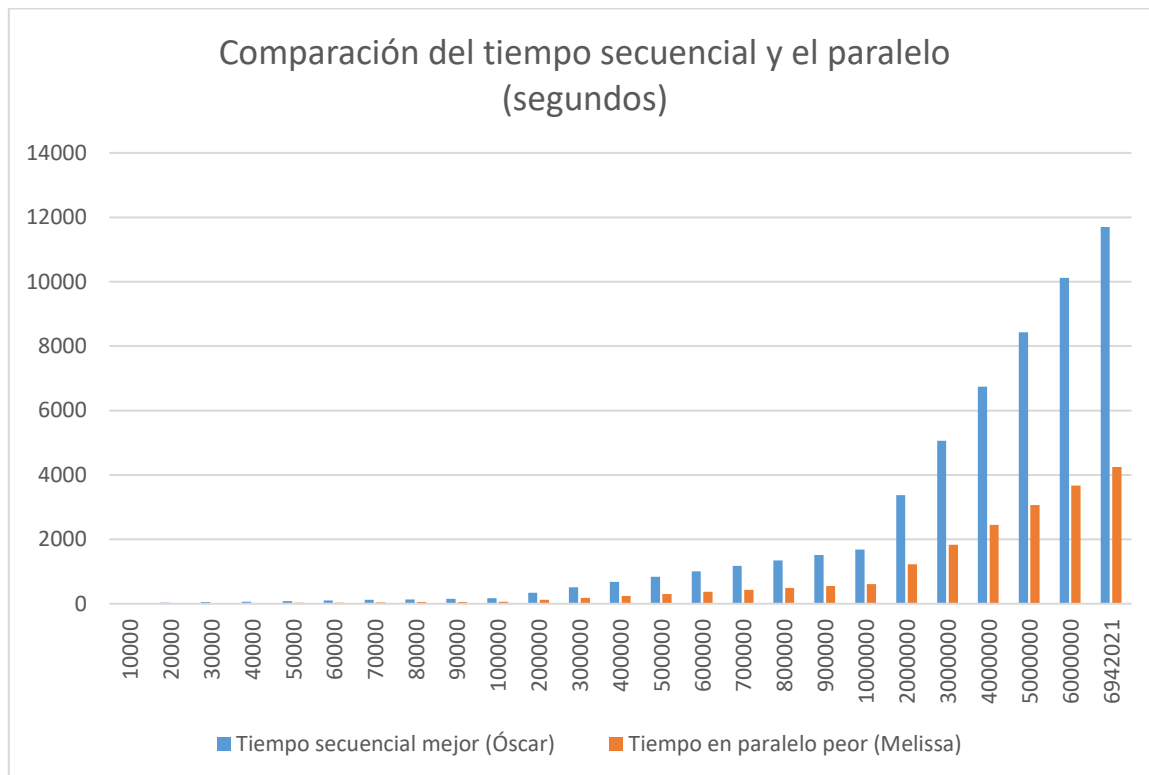
## Resultados de tiempo paralelo

Para el reporte de tiempo primeramente se tomarán todos los valores (véase anexo 11 para un mayor reporte de tiempos por integrante) que se ve reflejado en el siguiente gráfico:



La interpretación que el equipo le dio a los resultados del tiempo paralelo es que pudieron ocurrir algunos errores a lo largo del procedimiento, por ejemplo, la obtención del tiempo de complejidad lineal, que probablemente no era un tiempo escalable a tanto tiempo; por ello, el tiempo muy largo que Dariana reportó, causó que el algoritmo genético, al momento de escalar tanto los tiempos, le asignara tan pocos tweets al elevar la cantidad de tweets. También puede ser causa del algoritmo genético, aunque realiza buen trabajo para reducir los tiempos, – como más adelante se expondrá – es claro que un algoritmo genético encuentra buenas soluciones aunque no necesariamente las mejores.

Puede observarse que debido a la asignación de tiempos y del algoritmo, los tiempos que peor escalabilidad tuvieron fueron los de Melissa y Óscar, especialmente Melissa que para procesar su parte en paralelo del total de tweets, tomó una hora y 10 minutos, siendo este el peor de los tiempos para esta categoría, pero por otro lado, se mejoró en un 285% el tiempo base del procesamiento secuencial.



Puede concluirse que, como anteriormente se mencionó, se mejoró el tiempo en un 285%, es decir, que el algoritmo genético consiguió que la mejoría en tiempos por distribución de carga realmente estuviera funcionando.

# Aprendizaje Supervisado

Como ya sabemos inteligencia artificial es la combinación de algoritmos planteados con el propósito de crear máquinas que presenten las mismas capacidades que el ser humano. El machine learning es una rama de la computación de la IA que se basa en el entrenamiento de algoritmos de aprendizaje automático a partir de datos anteriores.

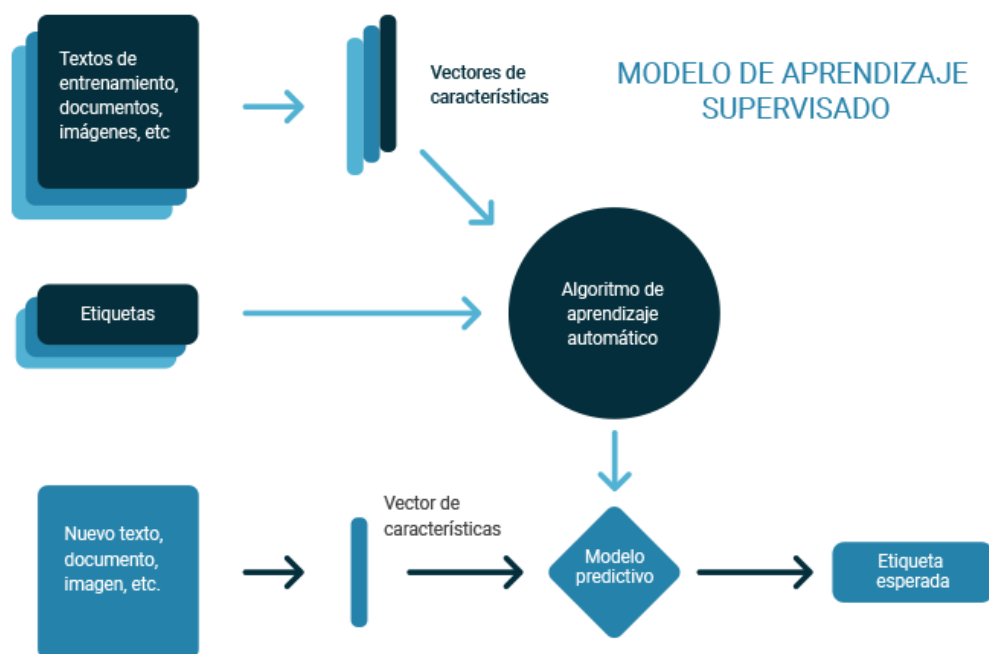
Esos tipos de aprendizajes se dividen en supervisados, no supervisados, semi-supervisado y por refuerzo; pero en esta ocasión nos centraremos en el aprendizaje supervisado.

El aprendizaje supervisado necesita conjuntos de datos etiquetados, es decir, le decimos al modelo qué es lo que queremos que aprenda. Dependiendo del tipo de etiqueta, dentro del aprendizaje supervisado existen dos tipos de modelos:

- **Los modelos de clasificación**, que producen como salida una etiqueta discreta, es decir, una etiqueta dentro de un conjunto finito de etiquetas posibles. A su vez, los modelos de clasificación pueden ser binarios si tenemos que predecir entre dos clases o etiquetas (enfermedad o no enfermedad, clasificación de correos electrónicos como “spam” o no “spam”) o multiclase, cuando se tiene que clasificar más de dos clases (clasificación de imágenes de animales, análisis de sentimientos, etc.).
- **Los modelos de regresión** producen como salida un valor real.

En los algoritmos de aprendizaje supervisado se genera un modelo predictivo, basado en datos de entrada y salida. La palabra clave “supervisado” viene de la idea de tener un conjunto de datos previamente etiquetado y clasificado, es decir, tener un conjunto de muestra, el cual ya se sabe a qué grupo, valor o categoría pertenecen los ejemplos. Con este grupo de datos que llamamos datos de entrenamiento, se realiza el ajuste al modelo inicial planteado. Es de esta forma como el algoritmo va “aprendiendo” a clasificar las muestras de entrada comparando el resultado del modelo, y la etiqueta real de la muestra, realizando las

compensaciones respectivas al modelo de acuerdo con cada error en la estimación del resultado.



Existen muchos algoritmos de aprendizaje supervisado que nos ayudan a clasificar, por ejemplo:

- Árboles de decisión
- Clasificación de Naïve Bayes
- Regresión por mínimos cuadrados
- Regresión Logística
- Support Vector Machines (SVM)
- Métodos "Ensemble" (Conjuntos de clasificadores).

Hablando un poco del modelo Support Vector Machine, es o, mejor dicho, son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik y su equipo en los laboratorios AT&T. Que funciona a grandes rasgos de la siguiente manera; Dado un conjunto de puntos, subconjunto de un conjunto mayor (espacio), en el que cada uno de ellos pertenece a una de dos posibles categorías, un algoritmo basado en SVM construye un modelo capaz de predecir si un punto nuevo (cuya categoría desconocemos) pertenece a una categoría o a la otra.



Como en la mayoría de los métodos de clasificación supervisada, los datos de entrada (los puntos) son vistos como un vector p-dimensional (una lista ordenada de p números).

La SVM busca un hiperplano que separe de forma óptima a los puntos de una clase de la de otra, que eventualmente han podido ser previamente proyectados a un espacio de dimensionalidad superior.

En ese concepto de "separación óptima" es donde reside la característica fundamental de las SVM: este tipo de algoritmos buscan el hiperplano que tenga la máxima distancia (margen) con los puntos que estén más cerca de él mismo. Por eso también a veces se les conoce a las SVM como clasificadores de margen máximo. De esta forma, los puntos del vector que son etiquetados con una categoría estarán a un lado del hiperplano y los casos que se encuentren en la otra categoría estarán al otro lado.

El link de los archivos “Team 10 Scrapping (Primera Limpieza).csv”, “Lemantizacion2.csv” y “Sentimiento.csv” se encuentra en este espacio:

[https://eduuuaa-my.sharepoint.com/:f/g/personal/al211800\\_edu\\_uuaa\\_mx/EpAen6GcultlmVLTuzvs5qMBbmNdO7D7\\_SKp3MCfT1PKDw?e=6sKbZW](https://eduuuaa-my.sharepoint.com/:f/g/personal/al211800_edu_uuaa_mx/EpAen6GcultlmVLTuzvs5qMBbmNdO7D7_SKp3MCfT1PKDw?e=6sKbZW).

El archivo CSV de tweets extraídos tiene una apariencia como la siguiente:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Unnamed: 0	Nombre de l	Nombre Visi	ID del Usuari	Descripci	Ubicaci	Tweet	Fecha de Creaci	Cantidad de	Idioma	En respuesta a	
2	0	0	æf'âxðÿ"â€	Lil_Nigger	1.43E+18	â€â€â€fâf% â%â,âiï%ï%ÿ	@Naushikan	23/08/2021 13:10		0	und	Naushikamaru2	
3	1	1	Croispin O'M	croispin	1.43E+18	Climate Fath	Mag Mell, Th	@talkRADIO	23/08/2021 13:09		0	en	talkRADIO
4	2	2	æf'âxðÿ"â€	Lil_Nigger	1.43E+18	â€â€â€fâf% â%â,âiï%ï%ÿ	@mame_chi	23/08/2021 13:08		0	ja	mame_chivi	
5	3	3	ï'€ï%œâššâ%	OnanismAsp	1.43E+18	(^_ ^)	RT @aya_is_	23/08/2021 13:08		0	ja		
6	4	4	2éšž3â"4ç	aya_is_kawa	1.43E+18	æ\$œéâ" ç šé âÿâ%â€ç,ââ.â	@OnanismA	23/08/2021 13:07		0	ja	OnanismAsp	
7	5	5	ðÿâ,âžâ"ðÿž	momiji_torit	1.43E+18	é"â€æ iç"ÿâ€é-œæš	@salmon_sp	23/08/2021 13:04		0	ja	salmon_spoon	
8	6	6	SJ.	leukhelurke	1.43E+18	Nee.	Komt de sco	23/08/2021 13:03		0	nl		
9	7	7	â%â,âiïâ	mame_chivi	1.43E+18		@Lil_Nigger	23/08/2021 13:02		0	ja	Lil_Nigger	
10	8	8	Liquor Sexy l	Am_MaXin	1.43E+18	Foodie	South Africa	RT	23/08/2021 13:02		0	en	
11	9	9	Sheryl obwa	ObwayaSher	1.43E+18	Just for funâ€	RT	23/08/2021 13:02		0	en		
12	10	10	minnie_gcwi		1.43E+18		Durban	I donâ€t bl	23/08/2021 13:02		0	en	
13	11	11	Nengified/li	lamideshaffy	1.43E+18	Wizkid fc/spartan/Titan/	RT	23/08/2021 13:01		0	en		
14	12	12	ðÿÿ,â€ðÿœ"lonely_playt		1.43E+18	ç°±æ%â€âf è±šç"	Nigger	23/08/2021 13:01		0	en		
15	13	13	iamnally_off	NallyNam	1.43E+18	Goal is to	Namibia	RT	23/08/2021 13:00		0	en	
16	14	14	Monkeymou	motorolaflip	1.43E+18	ï•œëµ-	@Ahmdyarr	23/08/2021 13:00		0	en	Ahmdyarr	
17	15	15	The Queen â	nahlalaa	1.43E+18	please donâ€t @ me un	RT @jordanc	23/08/2021 12:59		0	en		
18	16	16	â%â,âiïyana	IcyanaK	1.43E+18	ig: @alaskan	Fairbanks, al	RT @jordanc	23/08/2021 12:59		0	en	
19	17	17	nadiacantsw	nadiaoffthev	1.43E+18		she/her	RT @jordanc	23/08/2021 12:59		0	en	
20	18	18	Renesmee	wizardkeish	1.43E+18	If you donâ€t have a prc	RT @jordanc	23/08/2021 12:58		0	en		
21	19	19	Marc	_0720201219	1.43E+18	Add me on	ï„,ë ~ï•œï œ	RT @syvzwvi	23/08/2021 12:57		0	in	
22	20	20	Bucko.luvv.S	Jcantshoot	1.43E+18	Astros 72-50	Dallas, TX	@TheGoatW	23/08/2021 12:57		0	fr	TheGoatWiseman
23	21	21	Marc	_0720201219	1.43E+18	Add me on	ï„,ë ~ï•œï œ	RT @nami21	23/08/2021 12:57		0	in	
24	22	22	Liquorose's	cMrsHendrix	1.43E+18	Vegan. Accoi	#BBNaija		23/08/2021 12:57		11	en	
25	23	23	Marc	_0720201219	1.43E+18	Add me on	ï„,ë ~ï•œï œ	RT @shuuue	23/08/2021 12:57		0	in	
26	24	24	Sierraðÿÿ<	_Sierra1005	1.43E+18	USCB â€24ð	South Carolii	RT @jordanc	23/08/2021 12:56		0	en	
27	25	25	David Blaine	JohhnyTapia	1.43E+18	Zoe ideology	Cuba	RT @jordanc	23/08/2021 12:56		0	en	
28	26	26	lil book read	lextheloon	1.43E+18	Do you believe in the evo	i will never		23/08/2021 12:56		0	en	
29	27	27	carpediem	4lifeL	1.43E+18	Pending	The Hidden l	RT @jordanc	23/08/2021 12:56		0	en	

El archivo puede encontrarse adjunto en este entregable con el nombre “*Team 10 Scrapping (Primera Limpieza).csv*”.

Al hacer el proceso de lemantización se generó un nuevo archivo llamado “*Lemantizacion2.csv*” con una visualización como la de a continuación:

A	B	C	D	E	F	G	H	I	J	K	L
RT @ jordancoduu : @ poochivicious Did he just call me a nigger ?											
@ inteledits sir aap to nigger ho											
@ poochivicious Did he just call me a nigger ?											
@ KnockBrands Negrito Bimbo with a `` white afro kid " But now its call `` nito " cuss `` Negrito " be basically `` nigger " and `` Nito " Is the same but More frien											
@ Charlesrat74 If it be up to you negros wrestle would have all nigger champion											
@ 0X1HONG His as just look like he problematic .. maybe would call me a nigger or two											
Django : " .. Every single word that come out of Calvin Candie 's mouth be nothing but horseshit but he be right about one thing . I be that one nigger in ten t											
Precisava disso ?											
@ dking_go @ Tegs114 @ niccjennings @ BigBadCrabCake @ ballinhog @ itspreciouzzzz @ Big_Eti @ ChrisExcel102 @ EvrybodyHatesAP Stfu niggér											
@ DerrickNAACP Do you plan on go on after USTA.3 Black men be sue Black NYPD LT be call nigger by Black parking attend 2017 US Open Tennis a disgrace.NA											
nigger											
RT @ Phil_Lewis_ : Tarrant Alabama city council member Tommy Bryant have come under fire after he stand up and loudly ask " Do we have a ho...											
isa just call me a nigger SERIOUS NOT CLICKBAIT											
My quest4 # Justice4Vera & amp ; Title VII 1964 CRA victim be # Divine i.e . Unwed dead mom # EulaMaeRichardson firing by Mr Carter4 `` mention that nigger											
`` house nigger fuck allow to fuck leave "											
RT @ ChiefKeef : Fuck u nigger http : //t.co/pBhUZY1LiB											
FOOLISH NIGGER ? ? €\^ * \+ http : //t.co/ul47C6dUzV											
A fly roach just land on my front door . I ' m pretty sure bitch call me a nigger .											
Unprofessional nigger b.s exactly as I say it Automatically blame money and power											
NIGGER											
twitter term people for say  k\$ \$ but im get call a nigger and its cricket ? ?											
RT @ StanCollimore : Sharna Walker . Worcester . Racist . In custody as we speak . Sacked from a job . `` Black cunt " nigger " . http : //t.co/JGY...											
If you think this nigger will be the governor of California You 're on crack											
" Cut your hair . You should cut your hair . You know white people and corporate America cut your hair . Cut your hair . You should trim your beard . Cut your f											
Did you just call him a nigger ? http : //t.co/6iSti1IQbN											
RT @ Patrick42594797 : @ MatthewTPage @ USinNigeria @ AsstSecStateAF @ ECAatState US be now act on frustration @ StateDept Afgahnistan be a g...											
@ Puer_Hispano Creo que he interactuado con domicanos equivocados se parecen a los puerto rriqueños de Internet domados por los Yankis y el movimient											
@ N0IRXMINGS @ diiorhoseokk @ MOONSOFOHOPE @ shotasocean @ GLOHOB1 @ ELILICIOUSSS @ ksjkult @ horsoek @ makhylisgf @ s0ulscared @ billboardc											
@ MookieSZN50 NIGGER											
I love it dude It take the edge off of all the nigger bullshit we have to deal with !! I wan na see a war											
Take a hint nigger											
i could take my gun right now and shoot my jaw off and live but some fucking nigger in tarkov will grace my chin once with 9mm and i drop dead instantly											
@ reef1x_ @ Mongraal That ' s my nigger *											
Wendy stans come for crush like she didn ' t just call me a nigger last week y ' all sit down											
@ GuiltyG3kk Nigger activity http : //t.co/QRxZdFT18P											
I feel like Pulp Fiction have `` only " the one `` dead nigger storage " dialouge/scene but Reservoir Dogs have like 4 or 5 of them moment . Sheesh .											

Finalmente, se genera una evaluación de análisis de sentimientos en un archivo llamado “Sentimiento.csv” con la siguiente visualización:

	A	B	C	D	E
1	{'neg': 0.243	'neu': 0.567	'pos': 0.19	'compound': -0.3818}	
2	{'neg': 0.243	'neu': 0.567	'pos': 0.19	'compound': -0.3818}	
3	{'neg': 0.243	'neu': 0.567	'pos': 0.19	'compound': -0.3818}	
4	{'neg': 0.683	'neu': 0.317	'pos': 0.0	'compound': -0.6486}	
5	{'neg': 0.243	'neu': 0.567	'pos': 0.19	'compound': -0.3818}	
6	{'neg': 0.382	'neu': 0.412	'pos': 0.206	'compound': -0.6124}	
7	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
8	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
9	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
10	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
11	{'neg': 0.264	'neu': 0.736	'pos': 0.0	'compound': -0.6486}	
12	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
13	{'neg': 0.28	'neu': 0.72	'pos': 0.0	'compound': -0.6868}	
14	{'neg': 0.234	'neu': 0.602	'pos': 0.163	'compound': -0.5386}	
15	{'neg': 0.199	'neu': 0.694	'pos': 0.106	'compound': -0.4588}	
16	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
17	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
18	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
19	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
20	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
21	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
22	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
23	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
24	{'neg': 0.207	'neu': 0.728	'pos': 0.065	'compound': -0.5707}	
25	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
26	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
27	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
28	{'neg': 0.417	'neu': 0.583	'pos': 0.0	'compound': -0.6486}	
29	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
30	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
31	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
32	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
33	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
34	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	
35	{'neg': 0.0	'neu': 1.0	'pos': 0.0	'compound': 0.0}	

Con base en este archivo, se limpió el contenido no deseado, dejando únicamente la columna de compound que pondera la componente negativa, la neutra y la

positiva y da un resultado único para el tweet. Esta información fue la que se procesó en el código siguiente (véase el anexo 12 para revisar el código completo).

El link de los archivos “*RandomForest.py*” y “*Sentimiento.csv*” se encuentra en este espacio:

[https://eduuuaa-my.sharepoint.com/:f/g/personal/al211800\\_edu\\_uaa\\_mx/EtBbCLQY1gpCnV6rTvmgOiUBmHmjbcnf\\_A\\_zpolH1dg1cg?e=dJYTn2](https://eduuuaa-my.sharepoint.com/:f/g/personal/al211800_edu_uaa_mx/EtBbCLQY1gpCnV6rTvmgOiUBmHmjbcnf_A_zpolH1dg1cg?e=dJYTn2).

Primero se importaron las librerías:

```
### Importamos las librerías
import pandas as pd
import os
import numpy as np
import cv2
import matplotlib.pyplot as plt
import pickle
import random
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC

from sklearn.preprocessing import StandardScaler
from sklearn.naive_bayes import GaussianNB
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
```

Se cargó el archivo con el análisis de sentimientos, se definieron las categorías y se declaró la variable de datos:

```
### Cargamos los datos de análisis de sentimientos
datosRacismo = pd.read_csv("Sentimiento.csv")

### Declaramos las categorías
categories = ['Negativo', 'Neutro', 'Positivo']

### Declaramos la variable que mantendrá los datos a lo largo del procedimiento Random Forest
data = []
```

Se limpió la información para dejar únicamente la columna de compound del análisis de sentimientos:

```
### Se revisa por categoría cada imagen del directorio
for category in categories:
    label = categories.index(category)

    for i in datosRacismo.iloc[:,3]:
        numberData = i.replace(" \'compound\': ", "")
        numberData = numberData.replace("} ", "")
        data.append([float(numberData), round(float(numberData)) + 1])
```

Posteriormente se realiza la primera parte de análisis supervisado:

```
##### Se crea el documento data1 en el mismo directorio de archivos con el análisis Random Forest
pick_in = open('data1.pickle', 'wb')
pickle.dump(data, pick_in)
pick_in.close()

##### Se carga la información del archivo data1 en la variable data1
pick_in = open('data1.pickle', 'rb')
data1 = pickle.load(pick_in)
pick_in.close()

##### Se elige un individuo al azar
random.shuffle(data1)
features1 = []
label1 = []
```

Finalmente se realiza el procedimiento del algoritmo Random Forest y se muestran los resultados:

```
##### Se llena el arreglo de características de cada individuo
for feature, labelint in data1:
    features1.append([feature,0])
    label1.append(labelint)

##### Se realiza la clasificación a partir del entrenamiento usando 75% de la población
xtrain, xtest, ytrain, ytest = train_test_split(features1, label1, test_size = 0.20, random_state = 0)

sc = StandardScaler()

xtrain = sc.fit_transform(xtrain)
xtest = sc.fit_transform(xtest)

classifier = RandomForestClassifier(n_estimators = 12, random_state = 0)
classifier.fit(xtrain, ytrain)

ypred = classifier.predict(xtest)

print("Matriz de confusión:")
print(confusion_matrix(ytest, ypred))

print("Reporte de clasificación:")
print(classification_report(ytest, ypred))

print("Precisión:")
print(accuracy_score(ytest, ypred))
```

Los resultados que arroja el algoritmo en una ejecución son los siguientes (véase el anexo 13 para ver más ejecuciones):

```
Matriz de confusión:
[[421884      0      0]
 [    20 182436      0]
 [      0     33 24739]]
Reporte de clasificación:
              precision    recall  f1-score   support

      0       1.00      1.00      1.00     421884
      1       1.00      1.00      1.00     182456
      2       1.00      1.00      1.00      24772

 accuracy              1.00     629112
 macro avg           1.00      1.00      1.00     629112
weighted avg           1.00      1.00      1.00     629112

Precisión:
0.9999157542695101
```

La matriz de confusión está ordenada de izquierda a derecha (en columnas y de arriba hacia abajo en filas) por sentimiento de racismo negativo, neutro y positivo.

Como se puede observar, el algoritmo tiene una precisión muy grande, pues en la matriz de confusión, se observa que la diagonal, es decir, las predicciones acertadas componen a la mayoría, mientras que son pocos los tweets que llegan a ser mal validados.

b

a

## Conclusiones

**Andrea Melissa Almeida Ortega:** Para poder realizar el proyecto es imprescindible realizar una investigación adecuada acerca de nuestro tema, ya que necesitamos comprender muy bien que es el racismo y cómo lo demuestra la gente. Identificando las palabras principales que utilizan para ello, nos será más fácil clasificar nuestra base de datos en tweets negativos y positivos.

Como se pudo observar ya en la práctica de algoritmo genético simple enfocada en nuestro proyecto, nos damos cuenta que la optimización es algo esencial para el software y no es simplemente priorizar que realice las acciones más rápido, si no de entender lo que está haciendo el programa para poder repartir las distintas tareas y acciones para que se puedan ejecutar de la manera más eficaz posible.

Como se puede ver, es de suma importancia la implementación de una inteligencia artificial para clasificar cualquier cosa, en este caso, tweets. Cuando hablamos de proyectos grandes, ayuda a que empresas también sepan la opinión de personas a través de ellos.

**Joel Alejandro Espinoza Sánchez:** Me parece que como conclusión es temprano para evaluar los aprendizajes del proyecto, sin embargo, esta fase nos permite acercarnos a las herramientas que estaremos usando a lo largo del proyecto. Nos permite evaluar los atributos importantes que se necesitarán extraer de los tweets y será el primer cimiento para la inminente tabla de datos que realizaremos y con la que evaluaremos las próximas fases.

La implementación y uso del algoritmo genético fue muy importante para el proyecto de la materia y más importante con un planificador de carga y el procedimiento de lemantización. Con este programa se observó que la utilidad de un algoritmo genético puede ser muy importante en un procedimiento de optimización, y no necesariamente un algoritmo genético, pero para este proyecto fue fundamental.

Gracias a la actividad, nos permitimos explorar el planteamiento de un problema de aprendizaje desde la selección del algoritmo, ya que no todos los algoritmos están

diseñados para la solución de una misma tarea. Su diversidad permite la solución de problemas orientados a distintos tipos de datos, predicciones y objetivos y Random Forest era el idóneo.

Pude orientar personalmente un enfoque de investigación para proponer el mejor algoritmo con base en lo que deseábamos predecir y el conjunto de datos con lo que realizaríamos el procedimiento.

**Óscar Alonso Flores Fernández:** Al poder aprender a analizar la manera en que los usuarios de Twitter expresan sus opiniones relacionadas al tema del racismo, nos damos cuenta de que la mayoría de las personas se ve afectada de una u otra manera por distintos atributos, ya sean físicos o culturales, mostrándonos qué hay un sin fin de maneras despectivas de dirigirse a las personas, enfatizando una cualidad en específico.

Nuestro trabajo nos ha dado la oportunidad de identificar estas palabras y usarlas como fuente de investigación para profundizar cada vez más en cómo es el racismo desde el punto de vista de una de las redes sociales más importantes de la actualidad.

Al poder elaborar un algoritmo genético basándonos en lo que hemos estado elaborando podemos tener un esquema de trabajo mucho más directo y eficiente para poder llevar a cabo las distintas tareas, buscando siempre el poder cumplir con lo que el AG nos dicta y así optimizar el trabajo, la división de tareas, etc.

Poder trabajar de nuevo con Random Forest nos permite comprender el cómo hacer uso de diversas fuentes de recursos. En este caso el programa es sencillo y a una escala muy pequeña, pero escalar poco a poco con el fin de realmente mejorar la eficiencia en futuros proyectos de mayor tamaño será imperativo para ser capaces de cubrir las demandas de cada tarea llevada a cabo.

**Dariana Gómez Garza:** Realizar esta primera fase del proyecto nos hizo conocer un poco más sobre el tema del racismo y como es que se comportan las personas en redes sociales ante este, también como se expresan mediante tweets positivos o negativos sobre el tema elegido. Esta investigación es muy importante, ya que



nos ayuda a tomar en cuenta el contexto de los tweets y las palabras que se utilizan en ellos para poder clasificarlos.

Un avance también del proyecto fue la descarga de tweets que realizamos los integrantes del equipo.

En la elaboración de esta práctica implementamos para otro fin el algoritmo genético y otras herramientas útiles que nos ayudaron a observar las funciones de nuestro proyecto que tenemos en marcha de esta materia y lo útil que es para mejorar otros algoritmos. Las clases que tuvimos nos dieron una idea mejor de cómo realizar esta investigación. También nos dimos cuenta de cómo podríamos aplicar esto y que fuera funcional al mismo tiempo para nuestro proyecto en curso.

En este proyecto final pudimos agrandar el conocimiento que teníamos en aprendizaje supervisado, ya que nunca habíamos hecho una inteligencia artificial a base de tweets en algún curso.

Es interesante la amplia gama de opciones que hay para realizar este tipo de programas, por ejemplo, encontramos mucha información sobre Tensorflow y cómo era más sencillo clasificarlo así en lugar de realizarlo con el aprendizaje supervisado; pero el punto de este último parcial era tratar de realizarlo con este tema (Random Forest) en lugar de redes neuronales.

Me gustó mucho como quedó nuestro proyecto final y también era muy interesante como el algoritmo detectaba las emociones de los tweets y las clasificaba.

**Fernando Francisco González Arenas:** Al realizar las actividades de esta primera fase de recopilación de tweets e investigación sobre el tema del racismo, aprendí muchas cosas sobre la recopilación de información a través de las redes sociales y que es una gran fuente de información con la cual se pueden entrenar distintos modelos de inteligencia artificial para realizar tareas de reconocimiento de conductas humanas, lo cual es muy útil para identificar síntomas negativos como depresión, tristeza, racismo, discriminación, entre muchas otras conductas no tan positivas del ser humano.

Esta distribución de tweets permite utilidades en las que podemos agilizar el trabajo de las computadoras fácilmente. En este trabajo podemos observar cómo nuestras computadoras a partir de un programa en Python y un planificador de carga funcional podemos tener nuestras computadoras trabajando de manera acorde a su comportamiento.

El reconocimiento de emociones en los tweets por medio de aprendizaje supervisado es una técnica muy útil que tiene múltiples usos en la vida diaria de las personas alrededor del mundo en la actualidad, se puede usar para fines de seguridad, clasificación de grupos de población, detectar emociones y sentimientos de las personas, etc.

Con la realización de esta práctica investigamos formas de detectar emociones en los tweets y clasificarlos para saber si contenía palabras racistas o no, lo cual puede tener muchas aplicaciones prácticas en el futuro, incluso para futuros proyectos de los integrantes de este mismo equipo.

**Hiram Efraín Orocio García:** Al realizar esta primera fase del proyecto me ayudó a conocer más respecto al tema del racismo y la interacción en las redes sociales respecto a diversos temas y diversas opiniones aunque dichas opiniones pueden llegar a ser hirientes o afectar otras personas.

Un algoritmo genético es útil al momento de buscar soluciones por medio de exploración, así que aprender a implementar uno así sea en una aplicación básica, en un futuro lo podríamos implementar en problemas mayores que serán de gran ayuda.

La clasificación de emociones es útil al momento de hacer uso de recursos, datos, servicios, entre otras cosas así que aprender a implementar aun así sea en una aplicación básica como para saber qué opinan sobre un producto o en este caso saber que emociones transmiten las personas mediante las publicaciones, en un futuro lo podríamos implementar en problemas mayores donde el compartir estos recursos, serán de gran ayuda.

## Referencias

1. Bohórquez-Carvajal, J. D. (2020). *Razones y racismos. Antecedentes del determinismo biológico en el pensamiento ilustrado*. Utopia y Praxis Latinoamericana, 25, 430–446. <https://doi.org/10.5281/zenodo.4278398>.
2. Eubanks, O. (2020, 9 agosto). *Here are some commonly used terms that actually have racist origins*. ABC News. <https://abcnews.go.com/Politics/commonly-terms-racist-origins/story?id=71840410>.
3. Fukuzaki, R. (2021, 11 abril). *Everyday words and phrases that have racist connotations*. ABC7 Los Angeles. <https://abc7.com/racism-black-lives-matter-racist-words/6302853/>.
4. Isacc, B. (2020, 8 junio). *10 common phrases that are actually racist AF*. Upworthy. <https://www.upworthy.com/10-common-phrases-that-are-actually-racist-af>.
5. K. (2021, 10 julio). *43 Things a Karen Says*. Listcaboodle. <https://listcaboodle.com/43-things-a-karen-says/>.
6. Dara D. Mendez, Jewel Scott, Linda Adodoadji, Christina Toval, Monica McNeil, & Mahima Sindhu. (2021). *Racism as Public Health Crisis: Assessment and Review of Municipal Declarations and Resolutions Across the United States*. *Frontiers in Public Health*, 9. <https://doi.org/10.3389/fpubh.2021.686807>.
7. Thomas, M. D., Michaels, E. K., Darling-Hammond, S., Nguyen, T. T., Glymour, M. M., & Vittinghoff, E. (2020). *Whites' County-Level Racial Bias, COVID-19 Rates, and Racial Inequities in the United States*.
8. Hilda Beatriz Quintero Corrales, & Alexander Murillo Moreno. (2021). *El racismo como ideología y su negación en nuestras sociedades*. *Perspectivas Revista de Ciencias Sociales*, 11.
9. Luca Sebastiani. (2021). *Investigando los límites de la lucha legal contra el racismo: el marco español de antidiscriminación por origen racial o étnico*.

10. Tinoco, A. (2016, noviembre 23). *Hija de Eugenio Derbez sufre discriminación por hablar español*. La Opinión. <https://laopinion.com/2016/11/23/hija-de-eugenio-derbez-sufre-discriminacion-por-hablar-espanol/>.
11. Alejandro Tomás, & Hernán Dal Molin. (2021). *Racismo, subalternización política y derechos negados en América Latina*. Perspectivas Revista de Ciencias Sociales, 11.
12. *Understanding racism and how to spot it*. (s. f.). The Big Issues | ReachOut Australia. Recuperado 25 de agosto de 2021, de <https://au.reachout.com/articles/what-is-racism-and-how-to-spot-it>.
13. C. Darwin (1859). *On the Origin of Species by Means of Natural Selection*, Murray, London.
14. C. Reeves (1993). *Modern Heuristic Techniques for Combinatorial Problems*, Blackwell Scientific Publications.
15. D. E. Goldberg, Jr. R. Lingle (1985). *Alleles, loci and the traveling salesman problem*, en Proceedings of the First International Conference on Genetic Algorithms and Their Applications, 154-159.
16. Z. Michalewicz (1992). *Genetic Algorithms + Data Structures = Evolution Programs*, SpringerVerlag, Berlin Heidelberg.
17. Brad Miller, David Ranum. (2006). *Problem solving with algorithms and data structures using python*, Fraklin, Beedle & Associates.
18. Albert Y. Zomaya, Senior Member, IEEE, and Yee-Hwei Teh. (SEPTEMBER 2001). *Observations on Using Genetic Algorithms for Dynamic Load-Balancing*. IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, 12, 13. Octubre 10, 2021.
19. Lima, A. (2021). *Python: enfoques de lematización con ejemplos* – Acervo Lima. Retrieved 12 October 2021, from <https://es.acervolima.com/2021/02/09/python-enfoques-de-lematizacion-con-ejemplos/>

20. *Pandas desde cero* (2021). Retrieved 12 October 2021, from <https://www.youtube.com/watch?v=IhC01D6CbVU&list=PLgHCrivozIb0ULMKfJVV-rFdRG2OeEgfq&index=5>.
21. *Análisis de sentimiento. Qué es y cómo realizarlo* | QuestionPro. (2019). Retrieved 12 October 2021, from <https://www.questionpro.com/blog/es/herramienta-de-analisis-de-sentimientos/#:~:text=El%20an%C3%A1lisis%20de%20sentimiento%20es,sus%20actitudes%2C%20emociones%20y%20opiniones>.
22. *Análisis de sentimiento, ¿qué es, cómo funciona y para qué sirve?* (2017). Retrieved 12 October 2021, from <https://itelligent.es/es/analisis-de-sentimiento/>.
23. *Lematización* - Wikipedia, la enciclopedia libre. (2021). Retrieved 12 October 2021, from <https://es.wikipedia.org/wiki/Lematizaci%C3%B3n#:~:text=La%20lematizaci%C3%B3n%20es%20un%20proceso,flexionadas%20de%20una%20mism a%20palabra>.
24. *Lemantización basada en análisis supervisado del corpus* (2021). Retrieved 12 October 2021, from <https://users.dcc.uchile.cl/~abassi/ecos/lema.html>.
25. *Procesamiento del lenguaje natural* - EcuRed. (2021). Retrieved 12 October 2021, from [https://www.ecured.cu/Procesamiento\\_del\\_lenguaje\\_natural#:~:text=El%20%22Procesamiento%20del%20Lenguaje%20Natural,comprensi%C3%B3n%20autom%C3%A1tica%20del%20lenguaje%20natural](https://www.ecured.cu/Procesamiento_del_lenguaje_natural#:~:text=El%20%22Procesamiento%20del%20Lenguaje%20Natural,comprensi%C3%B3n%20autom%C3%A1tica%20del%20lenguaje%20natural).
26. *Procesamiento del lenguaje natural ¿qué es?* - IIC. (2017). Retrieved 12 October 2021, from <https://www.iic.uam.es/inteligencia/que-es-procesamiento-del-lenguaje-natural/>.
27. *Procesamiento del Lenguaje Natural con Machine Learning*. (2021). Retrieved 12 October 2021, from <https://www.encora.com/es/blog/procesamiento-del-lenguaje-natural-con-machine-learning>.
28. Igual, L. (2017). *Introduction to Data Science*. Barcelona: Springer.

29. Rebala, G. (2019). *An Introduction to Machine Learning*. California: Springer.
30. Sarkar, D. (2018). *Practical Machine Learning with Python*. Chicago: Apress.
31. Unpingco, J. (2019). *Python for Probability, Statistics and Machine Learning*. California: Springer.

## Anexos

### Anexo 1: Primera implementación en Python para la minería de la web.

```
import pandas as pd
import tweepy
k1 = 'U.U'
jorge = 'U-U'
k3 = 'UnU'
k4 = 'UwU'
aut = tweepy.OAuthHandler(k1, jorge)
aut.set_access_token(k3, k4)
api = tweepy.API(aut)
numt = 750
autor = []
contri = []
coord = []
cread = []
destroy = []
disp_text = []
enti = []
fav = []
fav_count = []
favd = []
ftxt = []
geo = []
iid = []
id_str = []
inscream = []
inrepstat = []
inrapstatstr = []
inrusid = []
inruseridstr = []
iscuotstat = []
lang = []
parse = []
parse_list = []
plays = []
retwitt = []
retwitt_count = []
retwitter = []
retwitts = []
source = []
source_ulr = []
truncater = []
```

```

user = []

twit = tweepy.Cursor(api.search, q="nigger",
tweet_mode="extended").items(numt)
for i in twit:
    autor.append(i.author)
    contri.append(i.contributors)
    coord.append(i.coordinates)
    cread.append(i.created_at)
    destroy.append(i.destroy)
    disp_text.append(i.display_text_range)
    enti.append(i.entities)
    fav.append(i.favorite)
    fav_count.append(i.favorite_count)
    favd.append(i.favorited)
    ftxt.append(i.full_text)
    geo.append(i.geo)
    iid.append(i.id)
    id_str.append(i.id_str)
    inscream.append(i.in_reply_to_screen_name)
    inrepstat.append(i.in_reply_to_status_id)
    inrapstatstr.append(i.in_reply_to_status_id_str)
    inrusid.append(i.in_reply_to_user_id)
    inruseridstr.append(i.in_reply_to_user_id_str)
    iscuotstat.append(i.is_quote_status)
    lang.append(i.lang)
    parse.append(i.parse)
    parse_list.append(i.parse_list)
    plays.append(i.place)
    retwitt.append(i.retweet)
    retwitt_count.append(i.retweet_count)
    retwitter.append(i.retweeted)
    retwitts.append(i.retweets)
    source.append(i.source)
    sourse_ulr.append(i.source_url)
    truncater.append(i.truncated)
    user.append(i.user)

df =
pd.DataFrame({'autor':autor, 'contri':contri, 'coord':coord, 'cread':
cread,
'destroy':destroy, 'disp_text':disp_text, 'enti':enti, 'fav':fav, 'fav
_count':fav_count, 'favd':favd, 'ftxt':ftxt, 'geo':geo, 'iid':iid, 'id
_str':id_str, 'inscream':inscream, 'inrepstat':inrepstat, 'inrapstatst
r':inrapstatstr, 'inrusid':inrusid, 'inruseridstr':inruseridstr, 'isc
uotstat':iscuotstat, 'lang':lang, 'parse':parse, 'parse_list':parse_l

```



```
ist,'plays':plays,'retwitt':retwitt,'retwitt_count':retwitt_count,  
'retwitter':retwitter,'retwitts':retwitts,'source':source,'source_  
ulr':source_ulr,'truncater':truncater,'user':user})  
df.to_csv()
```

## Anexo 2: Implementación formal del código en Python para la minería web.

```
### Uso de librerías
import pandas as pd
import tweepy

### Lectura de llaves
stream = open("keys.py")
read_file = stream.read()
exec(read_file)

### Conexión con la API
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)

### Definición de valores
num_tweets = 900
created_at = []
description = []
favorite_count = []
full_text = []
identifier = []
in_reply_to_screen_name = []
lang = []
location = []
name = []
screen_name = []
tweet = tweepy.Cursor(api.search, q="nigger",
tweet_mode="extended").items(num_tweets)

### Extracción de tweets
for i in tweet:
    created_at.append(i.created_at)
    description.append(i.author.description)
    favorite_count.append(i.favorite_count)
    full_text.append(i.full_text)
    identifier.append(i.id)
    lang.append(i.lang)
    location.append(i.author.location)
    in_reply_to_screen_name.append(i.in_reply_to_screen_name)
    name.append(i.author.name)
    screen_name.append(i.author.screen_name)
```

### Anexo 3: Código de server.py en Python.

```
### Conexión
import socket

s = socket.socket()
host = socket.gethostname()
port = 8080
s.bind((host,port))
s.listen(1)
print(host)
print("Esperando conexiones...")
conn, addr = s.accept()
print(addr + " se ha conectado")

### Transfiriendo archivos
filename = "Control.txt"
file = open(filename, 'rb')
file_data = file.read(1024)
conn.send(file_data)
print("Archivo enviado")
```

#### **Anexo 4: Código de client.py en Python.**

```
### Conexión
import socket

s = socket.socket()
print("Ingresa la dirección del host")
host = input()
port = 8080
s.connect((host,port))
print("Conectado")

### Transfiriendo archivos
filename = "Control.txt"
file = open(filename, 'wb')
file_data = s.recv(1024)
file.write(file_data)
file.close()
print("Recibido")
```

## **Anexo 5: Archivo de texto plano Control-Servidor.txt en su estado inicial.**

Estado: 0  
Cliente Melissa: 0  
Cliente Joel: 0  
Cliente Oscar: 0  
Cliente Dariana: 0  
Cliente Fernando: 0  
Cliente Hiram: 0

MELISSA  
Procesador (ultimo contacto): 0  
Historial del procesador:

Tweets enviados a procesar: 0  
Tweets ya procesados: 0  
Tiempo: 0

JOEL  
Procesador (ultimo contacto): 0  
Historial del procesador:

Tweets enviados a procesar: 0  
Tweets ya procesados: 0  
Tiempo: 0

OSCAR  
Procesador (ultimo contacto): 0  
Historial del procesador:

Tweets enviados a procesar: 0  
Tweets ya procesados: 0  
Tiempo: 0

DARIANA  
Procesador (ultimo contacto): 0  
Historial del procesador:

Tweets enviados a procesar: 0  
Tweets ya procesados: 0  
Tiempo: 0

FERNANDO  
Procesador (ultimo contacto): 0  
Historial del procesador:

Tweets enviados a procesar: 0  
Tweets ya procesados: 0  
Tiempo: 0

HIRAM  
Procesador (ultimo contacto): 0  
Historial del procesador:

Tweets enviados a procesar: 0  
Tweets ya procesados: 0  
Tiempo: 0

**Anexo 6: Archivo de texto plano Emision-Nombre.txt en su estado inicial (caso de ejemplo: Melissa).**

MELISSA

Procesador: 0

Tweets procesados: 0

Tiempo: 0

## **Anexo 7: Código con complejidad computacional lineal estándar en Python.**

```
import time

i = 0

inicio = time.time()
while i < 1000000:
    i = i + 0.01
final = time.time()
intervalo = final - inicio

print("Tiempo de ejecución: " + str(intervalo))
```



## Anexo 8: Evidencias de los resultados individuales al ejecutar el código del anexo 5.

Evidencia del tiempo de Melissa:

```
Tiempo de ejecución: 7.725000858306885

In [2]: runfile('C:/Users/mel_a/Documents/I.C.I/Python/untitled0.py',
wdir='C:/Users/mel_a/Documents/I.C.I/Python')
Tiempo de ejecución: 8.286704778671265

In [3]: runfile('C:/Users/mel_a/Documents/I.C.I/Python/untitled0.py',
wdir='C:/Users/mel_a/Documents/I.C.I/Python')
Tiempo de ejecución: 8.389048099517822

In [4]: runfile('C:/Users/mel_a/Documents/I.C.I/Python/untitled0.py',
wdir='C:/Users/mel_a/Documents/I.C.I/Python')
Tiempo de ejecución: 8.981545448303223

In [5]: runfile('C:/Users/mel_a/Documents/I.C.I/Python/untitled0.py',
wdir='C:/Users/mel_a/Documents/I.C.I/Python')
Tiempo de ejecución: 8.123982906341553

In [6]: runfile('C:/Users/mel_a/Documents/I.C.I/Python/untitled0.py',
wdir='C:/Users/mel_a/Documents/I.C.I/Python')
Tiempo de ejecución: 8.439027309417725

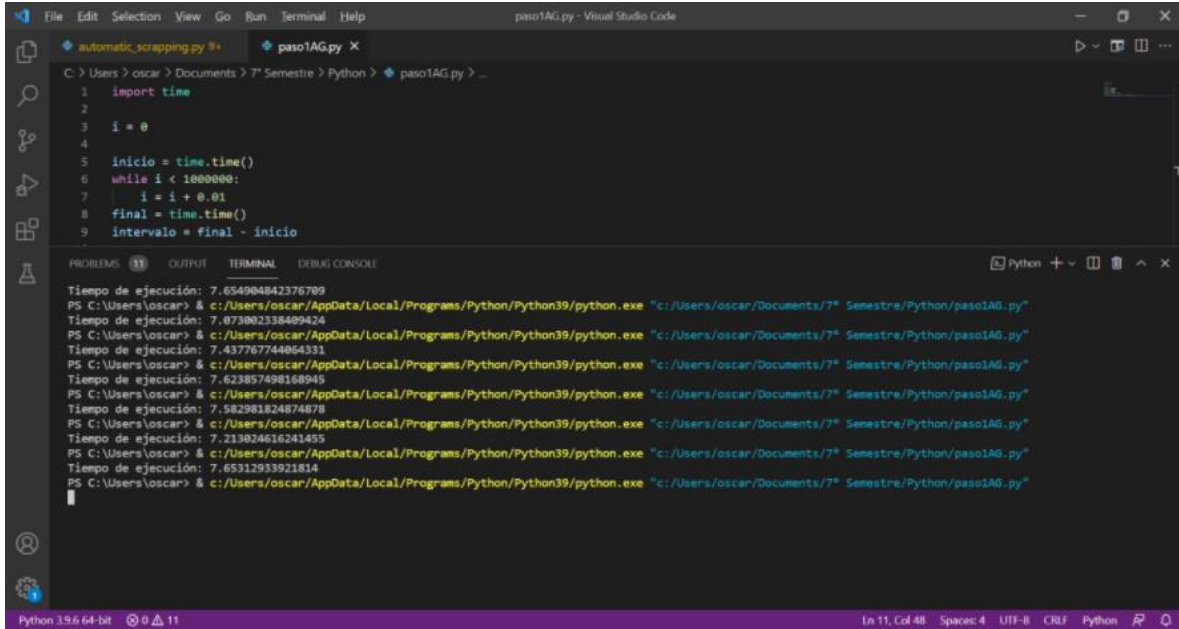
In [7]: runfile('C:/Users/mel_a/Documents/I.C.I/Python/untitled0.py',
wdir='C:/Users/mel_a/Documents/I.C.I/Python')
Tiempo de ejecución: 7.987653017044067

In [8]: runfile('C:/Users/mel_a/Documents/I.C.I/Python/untitled0.py',
wdir='C:/Users/mel_a/Documents/I.C.I/Python')
Tiempo de ejecución: 8.185094594955444

In [9]: runfile('C:/Users/mel_a/Documents/I.C.I/Python/untitled0.py',
wdir='C:/Users/mel_a/Documents/I.C.I/Python')
Tiempo de ejecución: 9.21445369720459

In [10]: runfile('C:/Users/mel_a/Documents/I.C.I/Python/untitled0.py',
wdir='C:/Users/mel_a/Documents/I.C.I/Python')
Tiempo de ejecución: 8.52103066444397
```

## Evidencia del tiempo de Óscar:



```
File Edit Selection View Go Run Terminal Help
paso1AG.py - Visual Studio Code

automatic_scraping.py 9+
paso1AG.py X

C:\Users\oscar> Documents\7° Semestre\Python> python paso1AG.py
1 import time
2
3 i = 0
4
5 inicio = time.time()
6 while i < 1000000:
7     i = i + 0.01
8     final = time.time()
9     intervalo = final - inicio

PROBLEMS OUTPUT TERMINAL DEBUG CONSOLE
Python + -

Tiempo de ejecución: 7.654904842376709
PS C:\Users\oscar> & c:/Users/oscar/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/oscar/Documents/7° Semestre/Python/paso1AG.py"
Tiempo de ejecución: 7.073002338409424
PS C:\Users\oscar> & c:/Users/oscar/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/oscar/Documents/7° Semestre/Python/paso1AG.py"
Tiempo de ejecución: 7.437767744064331
PS C:\Users\oscar> & c:/Users/oscar/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/oscar/Documents/7° Semestre/Python/paso1AG.py"
Tiempo de ejecución: 7.623857498168945
PS C:\Users\oscar> & c:/Users/oscar/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/oscar/Documents/7° Semestre/Python/paso1AG.py"
Tiempo de ejecución: 7.582981824874878
PS C:\Users\oscar> & c:/Users/oscar/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/oscar/Documents/7° Semestre/Python/paso1AG.py"
Tiempo de ejecución: 7.213024616241455
PS C:\Users\oscar> & c:/Users/oscar/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/oscar/Documents/7° Semestre/Python/paso1AG.py"
Tiempo de ejecución: 7.65312933921814
PS C:\Users\oscar> & c:/Users/oscar/AppData/Local/Programs/Python/Python39/python.exe "c:/Users/oscar/Documents/7° Semestre/Python/paso1AG.py"
```

## Evidencia del tiempo de Dariana:

```
PS C:\Users\DARIANA\Desktop> c:: cd 'c:\Users\DARIANA\Desktop'; & 'C:\Python39\python.exe' 'c:\Users\DARIANA\.vscode\extensions
Tiempo de ejecución: 90.12797212600708
PS C:\Users\DARIANA\Desktop> c:: cd 'c:\Users\DARIANA\Desktop'; & 'C:\Python39\python.exe' 'c:\Users\DARIANA\.vscode\extensions
\ms-python.python-2021.10.1317843341\pythonFiles\lib\python\debugpy\launcher' '49857' '--' 'c:\Users\DARIANA\Desktop\prueba.py'

Tiempo de ejecución: 89.6790201663971
PS C:\Users\DARIANA\Desktop> c:: cd 'c:\Users\DARIANA\Desktop'; & 'C:\Python39\python.exe' 'c:\Users\DARIANA\.vscode\extensions
\ms-python.python-2021.10.1317843341\pythonFiles\lib\python\debugpy\launcher' '49862' '--' 'c:\Users\DARIANA\Desktop\prueba.py'

Tiempo de ejecución: 90.48065257072449
PS C:\Users\DARIANA\Desktop> c:: cd 'c:\Users\DARIANA\Desktop'; & 'C:\Python39\python.exe' 'c:\Users\DARIANA\.vscode\extensions
\ms-python.python-2021.10.1317843341\pythonFiles\lib\python\debugpy\launcher' '49867' '--' 'c:\Users\DARIANA\Desktop\prueba.py'

Tiempo de ejecución: 93.18430972099304
PS C:\Users\DARIANA\Desktop> c:: cd 'c:\Users\DARIANA\Desktop'; & 'C:\Python39\python.exe' 'c:\Users\DARIANA\.vscode\extensions
\ms-python.python-2021.10.1317843341\pythonFiles\lib\python\debugpy\launcher' '49872' '--' 'c:\Users\DARIANA\Desktop\prueba.py'

Tiempo de ejecución: 90.65469336509705
```

Evidencia del tiempo de Hiram:

```
== RESTART: C:/Users/zente/AppData/Local,
Tiempo de ejecución: 22.436562538146973
>>>
== RESTART: C:/Users/zente/AppData/Local,
Tiempo de ejecución: 23.764434337615967
>>>
== RESTART: C:/Users/zente/AppData/Local,
Tiempo de ejecución: 20.85911536216736
>>>
== RESTART: C:/Users/zente/AppData/Local,
Tiempo de ejecución: 23.063973665237427
>>>
== RESTART: C:/Users/zente/AppData/Local,
Tiempo de ejecución: 25.14137578010559
>>>
== RESTART: C:/Users/zente/AppData/Local,
Tiempo de ejecución: 23.372560024261475
>>>
== RESTART: C:/Users/zente/AppData/Local,
Tiempo de ejecución: 22.626899003982544
>>>
== RESTART: C:/Users/zente/AppData/Local,
Tiempo de ejecución: 22.254277229309082
>>>
== RESTART: C:/Users/zente/AppData/Local,
Tiempo de ejecución: 23.449257373809814
>>>
== RESTART: C:/Users/zente/AppData/Local,
Tiempo de ejecución: 23.12931776046753
>>> |
```

## Anexo 9: Código del algoritmo genético en Python.

```
### Portada
'''
        Centro de Ciencias Básicas
    Departamento de Ciencias de la Computación
        Metaheurísticas I
            7º "A"

        Segunda Evaluación Parcial

        Profesor: Francisco Javier Luna Rosas

        Alumnos:
            Almeida Ortega Andrea Melissa
            Espinoza Sánchez Joel Alejandro
            Flores Fernández Óscar Alonso
            Gómez Garza Dariana
            González Arenas Fernando Francisco
            Orocio García Hiram Efraín

    Fecha de Entrega: Aguascalientes, Ags., 31 de octubre de 2021
'''

### Descripción

### Importación de librerías
import numpy as np
import random

### Funciones de apoyo
# Función FirstSample
def FirstSample(qe,qi,total_tweets):
    sample = []
    for i in range(qi):
        sample.append([])
        remaining_tweets = total_tweets
        for j in range(qe):
            sample[i].append(random.randint(0, remaining_tweets))
            remaining_tweets = remaining_tweets - sample[i][j]
        if remaining_tweets != 0:
            assignment = random.randint(1,6)
            sample[i][assignment - 1] = sample[i][assignment - 1] + remaining_tweets
    return sample

# Función PrintSample
def PrintSample(qi,sample):
    for i in range(qi):
        try:
            print(str(sample[0][i]) + " = " + str(sum(sample[0][i])) + " \t\t " +
str(sample[1][i]) + " \t " + str(sample[2][i]))
        except:
            try:
                print(str(sample[0][i]) + " = " + str(sum(sample[0][i])) + " \t\t " +
str(sample[1][i]) + " \t " + str(sample[2]))
            except:
                print(str(sample[0][i]) + " = " + str(sum(sample[0][i])) + " \t\t " +
str(sample[1]) + " \t " + str(sample[2]))
    return
```

```

# Función evaluate
def evaluate(qi,sample,power_process):
    evaluation = []
    for i in range(qi):
        function = -(np.abs(sample[0][i][0] - 1752929) + np.abs(sample[0][i][1] - 1752929)
+ np.abs(sample[0][i][2] - 2188458) + np.abs(sample[0][i][3] - 179704) +
np.abs(sample[0][i][4] - 1013759) + np.abs(sample[0][i][5] - 709446)) + 34710105

        print("Función objetivo " + str(i))
        print(function)

        evaluation.append(function)
    return evaluation

# Función evaluateA
def evaluateA(qi,sample):
    evaluation = []
    evaluation.append(sample[1][0])
    i = 1
    while i < qi:
        evaluation.append(evaluation[i - 1] + sample[1][i])
        i = i + 1
    return evaluation

# Función Sorting
def Sorting(sample, qi):
    i = 0
    while i < qi - 1:
        if sample[1][i] < sample[1][i + 1]:
            for j in range(3):
                aux = sample[j][i]
                sample[j][i] = sample[j][i + 1]
                sample[j][i + 1] = aux
            i = 0
            continue
        i = i + 1
    return sample

# Función nullNewSample
def newSampleCreator(qi):
    nullCreator = []
    for i in range(qi):
        nullCreator.append(0)
    nullNewSample = []
    nullNewSample.append(nullCreator[:])
    nullNewSample.append(nullCreator[:])
    nullNewSample.append(nullCreator[:])
    return nullNewSample

#%% Calibrar variables
total_tweets = 6942021
qe = 6 # Inamovible
qi = 60 # Movable
pc = 75 # Movable
pm = 40 # Movable
qelitism = 5 # Movable
qg = 10
sample = [[], [], []]
power_process = [9.311, 14.8685, 7.458, 90.8248, 16.1, 23.006] # Orden alfabético: Melissa,
Joel, Óscar, Dariana, Fernando, Hiram
ig = 0

```

```

%% AG: Creación de la población aleatoria
sample[0] = FirstSample(qe,qi,total_tweets)
PrintSample(qi, sample)

%% AG: Procedimiento de cada generación
while ig < qg:
    print("Población " + str(ig))
    print("")

    # Evaluamos a todas las muestras
    sample[1] = evaluate(qi,sample,power_process)

    print("Población " + str(ig) + " al ser evaluada")
    PrintSample(qi,sample)
    print("")

    # Evaluamos la suma acumulada
    sample[2] = evaluateA(qi,sample)

    print("Población " + str(ig) + " al ser evaluada con acumulación")
    PrintSample(qi,sample)
    print("")

    # Ordenamos por valor de f(x) es decir, por sample[1]
    sample = Sorting(sample, qi)

    print("Población " + str(ig) + " al ser ordenada")
    PrintSample(qi,sample)
    print("")

    # Reevaluamos la suma acumulada
    sample[2] = evaluateA(qi,sample)

    print("Población " + str(ig) + " al ser reevaluada")
    PrintSample(qi,sample)
    print("")

    # Creamos una nueva muestra vacía
    newSample = newSampleCreator(qi)

    print("Nueva muestra " + str(ig + 1) + " vacía")
    print(newSample)
    print("")

    print("Se aplicará elitismo")
    x = input()

    # Se aplica elitismo
    ii = 0
    while ii < qelitism:
        newSample[0][ii] = sample[0][ii]
        #newSample[1][ii] = sample[1][ii]
        #newSample[2][ii] = sample[2][ii]
        ii = ii + 1

    print("Nueva muestra " + str(ig + 1) + " después de elitismo")
    for i in range(qi):
        try:
            print(str(newSample[0][i]) + " = " + str(sum(newSample[0][i])))
        except:

```

```

        pass
    print("")
    print("Se aplicará cruzamiento")
    x = input()

    # Se aplicará cruzamiento
    while ii < qi:
        # Seleccionamos por ruleta
        # Padre 1
        random_pick = random.uniform(0, sample[2][qi - 1])

        for ic in range(qi):
            if random_pick < sample[2][ic]:
                c1 = sample[0][ic]
                print("Se elige como padre 1 el individuo " + str(ic) + "(random_pick = "
+ str(random_pick) + ")")
                break

        # Padre 2
        random_pick = random.uniform(0, sample[2][qi - 1])

        for ic in range(qi):
            if random_pick < sample[2][ic]:
                c2 = sample[0][ic]
                print("Se elige como padre 2 el individuo " + str(ic) + "(random_pick = "
+ str(random_pick) + ")")
                break

        # Se evalúa si se aplicará cruzamiento
        random_num = random.randint(0,99)
        if random_num < pc:
            print("Los cromosomas " + str(ii) + " y " + str(ii + 1) + " han activado
cruzamiento (Probabilidad: " + str(random_num) + " < " + str(pc) + ")")
            print("")

            # Mezclamos los 6 números de una tupla
            processors = [0, 1, 2, 3, 4, 5]
            random.shuffle(processors) # Se mezclarán y se harán sumas balanceadas de par
en par

            print("Tupla mezclada: " + str(processors))
            print("")

            print("Valores antes del cruzamiento: ")
            print(c1)
            print(c2)
            print("")

            for i in [0, 2, 4]:
                # Sumas equilibradas
                min_value = min(c1[processors[i]], c1[processors[i + 1]],
c2[processors[i]], c2[processors[i + 1]])
                operation = random.randint(0, 1) # Elige la operación con la que empezará
la suma equilibrada
                addition = random.randint(0, min_value)

                print("Tuplas " + str(processors[i] + 1) + " y " + str(processors[i + 1] +
1))

                print("min_value = " + str(min_value))
                print("addition = " + str(addition))
                print("")

```

```

        if operation == 0:
            c2[processors[i + 1]] = c2[processors[i + 1]] - addition
            c1[processors[i + 1]] = c1[processors[i + 1]] + addition
            c1[processors[i]] = c1[processors[i]] - addition
            c2[processors[i]] = c2[processors[i]] + addition
        else:
            c2[processors[i + 1]] = c2[processors[i + 1]] + addition
            c1[processors[i + 1]] = c1[processors[i + 1]] - addition
            c1[processors[i]] = c1[processors[i]] + addition
            c2[processors[i]] = c2[processors[i]] - addition

    print("Valores después del cruzamiento: ")
    print(c1)
    print(c2)
    print("")

    else:
        print("Los cromosomas " + str(ii) + " y " + str(ii + 1) + " no han activado
cruzamiento (Probabilidad: !" + str(random_num) + " < " + str(pc) + "!)")

        newSample[0][ii] = c1
        try:
            newSample[0][ii + 1] = c2
        except:
            pass

        ii = ii + 2

    print("Nueva muestra " + str(ig + 1) + " después del cruzamiento")
    for i in range(qi):
        try:
            print(str(newSample[0][i]) + " = " + str(sum(newSample[0][i])))
        except:
            pass
    print("")
    print("Se aplicará mutación")
    x = input()

    # Se aplica mutación
    ii = 0
    while ii < qi:
        # Se evalúa si se aplicará mutación
        random_num = random.randint(0,99)
        if random_num < pm:
            print("El cromosoma " + str(ii) + " ha activado mutación (Probabilidad: " +
str(random_num) + " < " + str(pm) + ")")
            print("")

            processors = [0, 1, 2, 3, 4, 5]
            random.shuffle(processors) # Se mezclarán y se hará una sumas balanceadas
cíclica

            print("Tupla mezclada: " + str(processors))
            print("")

            print("Valores antes de la mutación: ")
            print(newSample[0][ii])
            print("")

            min_value = min(newSample[0][ii])

```



```

        operation = random.randint(0, 1) # Elige la operación con la que empezará la
suma equilibrada
        addition = random.randint(0, min_value)

        print("Tupla " + str(processors))
        print("min_value = " + str(min_value))
        print("addition = " + str(addition))
        print("")

        if operation == 0:
            newSample[0][ii][processors[0]] = newSample[0][ii][processors[0]] -
addition
            newSample[0][ii][processors[1]] = newSample[0][ii][processors[1]] +
addition
            newSample[0][ii][processors[2]] = newSample[0][ii][processors[2]] -
addition
            newSample[0][ii][processors[3]] = newSample[0][ii][processors[3]] +
addition
            newSample[0][ii][processors[4]] = newSample[0][ii][processors[4]] -
addition
            newSample[0][ii][processors[5]] = newSample[0][ii][processors[5]] +
addition
        else:
            newSample[0][ii][processors[0]] = newSample[0][ii][processors[0]] +
addition
            newSample[0][ii][processors[1]] = newSample[0][ii][processors[1]] -
addition
            newSample[0][ii][processors[2]] = newSample[0][ii][processors[2]] +
addition
            newSample[0][ii][processors[3]] = newSample[0][ii][processors[3]] -
addition
            newSample[0][ii][processors[4]] = newSample[0][ii][processors[4]] +
addition
            newSample[0][ii][processors[5]] = newSample[0][ii][processors[5]] -
addition

        print("Valores después de la mutación: ")
        print(newSample[0][ii])
        print("")

        else:
            print("El cromosoma " + str(ii) + " no ha activado mutación (Probabilidad: !"
+ str(random_num) + " < " + str(pm) + "!)")
            print("")
            ii = ii + 1

        print("Nueva muestra " + str(ig + 1) + " después de la mutación")
        for i in range(qi):
            try:
                print(str(newSample[0][i]) + " = " + str(sum(newSample[0][i])))
            except:
                pass
        print("")
        print("Se tratará una nueva generación")
        x = input()

        sample = newSample
        ig = ig + 1

```

**Anexo 10: Tabla de tiempo secuencial (en segundos) por integrante según la cantidad de tweets.**

Tweets	Tiempo de Melissa	Tiempo de Joel	Tiempo de Óscar	Tiempo de Dariana	Tiempo de Fernando	Tiempo de Hiram
10,000	24.237	29.970	16.859	68.546	33.555	37.527
20,000	48.473	59.940	33.718	137.092	67.109	75.054
30,000	72.710	89.910	50.577	205.638	100.664	112.581
40,000	96.947	119.879	67.436	274.184	134.218	150.107
50,000	121.183	149.849	84.295	342.730	167.773	187.634
60,000	145.420	179.819	101.153	411.276	201.328	225.161
70,000	169.657	209.789	118.012	479.822	234.882	262.688
80,000	193.894	239.759	134.871	548.369	268.437	300.215
90,000	218.130	269.729	151.730	616.915	301.992	337.742
100,000	242.367	299.699	168.589	685.461	335.546	375.269
200,000	484.734	599.397	337.178	1370.921	671.092	750.537
300,000	727.101	899.096	505.767	2056.382	1006.638	1125.806
400,000	969.468	1198.794	674.356	2741.843	1342.185	1501.074
500,000	1211.835	1498.493	842.945	3427.303	1677.731	1876.343
600,000	1454.202	1798.191	1011.534	4112.764	2013.277	2251.612
700,000	1696.569	2097.890	1180.123	4798.225	2348.823	2626.880
800,000	1938.936	2397.589	1348.713	5483.685	2684.369	3002.149
900,000	2181.302	2697.287	1517.302	6169.146	3019.915	3377.417
1,000,000	2423.669	2996.986	1685.891	6854.607	3355.461	3752.686
2,000,000	4847.339	5993.971	3371.781	13709.213	6710.923	7505.372
3,000,000	7271.008	8990.957	5057.672	20563.820	10066.384	11258.058
4,000,000	9694.678	11987.943	6743.563	27418.426	13421.846	15010.744
5,000,000	12118.347	14984.928	8429.453	34273.033	16777.307	18763.430
6,000,000	14542.016	17981.914	10115.344	41127.639	20132.769	22516.116
6,942,021	16825.164	20805.137	11703.489	47584.823	23293.684	26051.225

# **Anexo 11: Tabla de tiempo en paralelo (en segundos) por integrante según la cantidad de tweets.**

Es necesario tomar en cuenta la tabla 1 que aparece en el apartado de resultados donde se dividen los tweets, pues el tiempo reportado aquí es en función de los tweets obtenidos en dicha tabla.

Tweets	Tiempo de Melissa	Tiempo de Joel	Tiempo de Óscar	Tiempo de Dariana	Tiempo de Fernando	Tiempo de Hiram
10,000	5.611	5.368	5.410	1.391	4.141	4.683
20,000	12.099	9.659	10.533	3.928	9.566	7.929
30,000	18.127	14.508	15.977	5.196	14.496	11.727
40,000	24.106	19.417	21.306	6.909	19.267	15.712
50,000	30.754	23.505	26.554	8.945	24.710	18.944
60,000	37.235	27.797	32.094	9.809	29.558	22.835
70,000	42.846	33.168	37.423	11.523	33.944	27.245
80,000	48.985	37.882	42.198	15.498	39.729	30.089
90,000	55.187	42.518	47.921	15.608	44.383	34.202
100,000	61.222	47.364	53.480	16.396	49.215	38.109
200,000	122.732	94.372	106.452	34.849	97.932	76.776
300,000	183.215	142.651	159.699	52.184	147.120	114.919
400,000	244.524	189.904	212.818	70.047	197.009	152.276
500,000	305.756	237.256	265.776	88.562	244.274	192.569
600,000	366.914	284.699	318.876	106.486	294.110	229.991
700,000	428.282	331.880	372.189	123.575	342.338	269.199
800,000	489.290	379.508	425.327	141.356	392.676	306.054
900,000	551.259	425.950	478.377	159.486	440.602	345.611
1,000,000	612.381	473.437	531.332	178.014	490.535	382.917
2,000,000	1224.106	947.683	1063.086	354.315	980.204	766.805
3,000,000	1835.881	1421.866	1594.450	532.226	1469.702	1150.870
4,000,000	2447.846	1895.518	2125.664	711.385	1959.730	1534.364
5,000,000	3059.781	2369.806	2657.571	886.376	2450.087	1917.480
6,000,000	3672.528	2842.785	3188.916	1064.342	2939.961	2301.140
6,942,021	4248.520	3289.863	3689.501	1231.800	3401.629	2662.332

## Anexo 12: Código de Machine Learning con Random Forest en Python:

```
### Presentación
'''
    Universidad Autónoma de Aguascalientes

        Centro de Ciencias Básicas
    Departamento de Ciencias de la Computación
        Metaheurísticas I
            7º "A"

        Actividad 5.01

    Profesor: Francisco Javier Luna Rosas

    Alumnos:
        Almeida Ortega Andrea Melissa
        Espinoza Sánchez Joel Alejandro
        Flores Fernández Óscar Alonso
        Gómez Garza Dariana
        González Arenas Fernando Francisco
        Orocio García Hiram Efraín

    Fecha de Entrega: 19 de noviembre del 2021

    Descripción: Machine Learning (Random Forest como Aprendizaje Supervisado)
    para análisis de sentimientos
    '''

### Importamos las librerías
import pandas as pd
import os
import numpy as np
import cv2
import matplotlib.pyplot as plt
import pickle
import random
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC

from sklearn.preprocessing import StandardScaler
from sklearn.naive_bayes import GaussianNB
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix,
accuracy_score

### Cargamos los datos de análisis de sentimientos
datosRacismo = pd.read_csv("Sentimiento.csv")

### Declaramos las categorías
```

```

categories = ['Negativo', 'Neutro', 'Positivo']

#%% Declaramos la variable que mantendrá los datos a lo largo del
procedimiento Random Forest
data = []

#%% Se revisa por categoría cada imagen del directorio
for category in categories:
    label = categories.index(category)

    for i in datosRacismo.iloc[:,3]:
        numberData = i.replace(" \'compound\': ", "")
        numberData = numberData.replace("} ", "")
        data.append([float(numberData), round(float(numberData)) + 1])

#%% Se crea el documento data1 en el mismo directorio de archivos con el
análisis Random Forest
pick_in = open('data1.pickle', 'wb')
pickle.dump(data, pick_in)
pick_in.close()

#%% Se carga la información del archivo data1 en la variable data1
pick_in = open('data1.pickle', 'rb')
data1 = pickle.load(pick_in)
pick_in.close()

#%% Se elige un individuo al azar
random.shuffle(data1)
features1 = []
label1 = []

#%% Se llena el arreglo de características de cada individuo
for feature, labelint in data1:
    features1.append([feature,0])
    label1.append(labelint)

#%% Se realiza la clasificación a partir del entrenamiento usando 75% de
la población
xtrain, xtest, ytrain, ytest = train_test_split(features1, label1,
test_size = 0.20, random_state = 0)

sc = StandardScaler()

xtrain = sc.fit_transform(xtrain)
xtest = sc.fit_transform(xtest)

classifier = RandomForestClassifier(n_estimators = 12, random_state = 0)
classifier.fit(xtrain, ytrain)

ypred = classifier.predict(xtest)

```

```
print("Matriz de confusión:")  
print(confusion_matrix(ytest, ypred))  
  
print("Reporte de clasificación:")  
print(classification_report(ytest, ypred))  
  
print("Precisión:")  
print(accuracy_score(ytest, ypred))
```

### Anexo 13: Ocho ejecuciones del código de Random Forest:

Ejecución 1:

```
Matriz de confusión:
[[421884      0      0]
 [      20 182436      0]
 [      0      33 24739]]
Reporte de clasificación:
              precision    recall  f1-score   support

      0       1.00      1.00      1.00     421884
      1       1.00      1.00      1.00     182456
      2       1.00      1.00      1.00      24772

 accuracy          1.00
macro avg          1.00      1.00      1.00     629112
weighted avg       1.00      1.00      1.00     629112

Precisión:
0.9999157542695101
```

Ejecución 2:

```
Matriz de confusión:
[[421621      0      0]
 [      0 182854      0]
 [      0      0 24637]]
Reporte de clasificación:
              precision    recall  f1-score   support

      0       1.00      1.00      1.00     421621
      1       1.00      1.00      1.00     182854
      2       1.00      1.00      1.00      24637

 accuracy          1.00
macro avg          1.00      1.00      1.00     629112
weighted avg       1.00      1.00      1.00     629112

Precisión:
1.0
```

Ejecución 3:

```
Matriz de confusión:
[[422121      0      0]
 [    12 182175      0]
 [      0     34 24770]]
Reporte de clasificación:
              precision    recall  f1-score   support

      0       1.00      1.00      1.00     422121
      1       1.00      1.00      1.00     182187
      2       1.00      1.00      1.00      24804

   accuracy                1.00     629112
  macro avg       1.00      1.00      1.00     629112
weighted avg       1.00      1.00      1.00     629112

Precisión:
0.9999268810641031
```

Ejecución 4:

```
Matriz de confusión:
[[421871     38      0]
 [      0 182654     31]
 [      0      0 24518]]
Reporte de clasificación:
              precision    recall  f1-score   support

      0       1.00      1.00      1.00     421909
      1       1.00      1.00      1.00     182685
      2       1.00      1.00      1.00      24518

   accuracy                1.00     629112
  macro avg       1.00      1.00      1.00     629112
weighted avg       1.00      1.00      1.00     629112

Precisión:
0.9998903215961545
```

Ejecución 5:



```

Matriz de confusión:
[[421815      0      0]
 [      21 182513      0]
 [      0      30 24733]]
Reporte de clasificación:
              precision    recall  f1-score   support

     0       1.00      1.00      1.00     421815
     1       1.00      1.00      1.00     182534
     2       1.00      1.00      1.00      24763

 accuracy          1.00
 macro avg          1.00
weighted avg          1.00

Precisión:
0.9999189333536794

```

Ejecución 6:

```

Matriz de confusión:
[[422502      43      0]
 [      0 181813      34]
 [      0      0 24720]]
Reporte de clasificación:
              precision    recall  f1-score   support

     0       1.00      1.00      1.00     422545
     1       1.00      1.00      1.00     181847
     2       1.00      1.00      1.00      24720

 accuracy          1.00
 macro avg          1.00
weighted avg          1.00

Precisión:
0.9998776052594769

```

Ejecución 7:

```

Matriz de confusión:
[[422274      37      0]
 [      0 181887      0]
 [      0      36 24878]]
Reporte de clasificación:

```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	422311
1	1.00	1.00	1.00	181887
2	1.00	1.00	1.00	24914
accuracy			1.00	629112
macro avg	1.00	1.00	1.00	629112
weighted avg	1.00	1.00	1.00	629112

```

Precisión:
0.9998839634278157

```

Ejecución 8:

```

Matriz de confusión:
[[423196      51      0]
 [      0 181283     110]
 [      0      0 24472]]
Reporte de clasificación:

```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	423247
1	1.00	1.00	1.00	181393
2	1.00	1.00	1.00	24472
accuracy			1.00	629112
macro avg	1.00	1.00	1.00	629112
weighted avg	1.00	1.00	1.00	629112

```

Precisión:
0.9997440837243606

```