





TECNOLOGICO NACIONAL DE MEXICO

INSTITUTO TECNOLOGICO DE MORELIA

"José María Morelos y Pavón"

Búsqueda Y Almacenamiento De Datos Ferreira Escutia Rogelio

Proyecto BAD Political Analytics

Carlos Jahir Castro Cázares Giovanni Hasid Martínez Reséndiz Jaime Isai Velazquez Aguilar

Semestre Sep - Ene 2020 - 2021 02 de febrero de 2021





Descripción del proyecto.

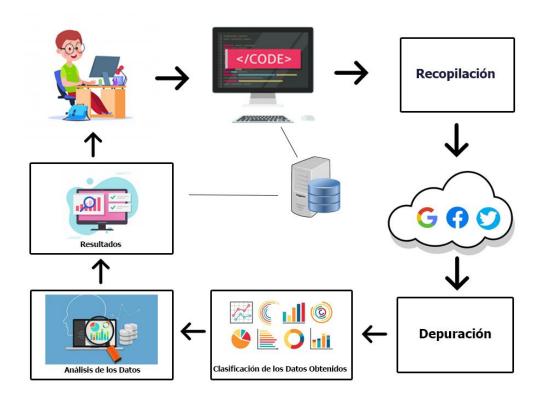
Political Analytics se desarrolla en el área de la política, en las elecciones por venir para predecir quien será el gobernador de Michoacán, el objetivo es obtener información relevante de los candidatos electorales a través de las redes sociales, con ayuda de Apis y web scrapping identificando cuantas personas lo siguen.

La intención es mostrar de manera gráfica como estos resultados van cambiando conforme a los días y poder hacer una correlación de los datos y por lo tanto un pronóstico con mayor acierto.

Diagrama a bloques del sistema.

Nuestro sistema usa un modelo cliente – servidor, donde el cliente pide datos al servidory este busca la informacion en internet, la depura y analiza esta misma para guardarla en la base de datos y para mostrar los resultados el cliente pide al servidor los datos guardados en la base de datos y los regresa para visualizarlos en la web.

Diagrama de Bloques



Estructura de las bases de datos y tablas empleadas.

Las tablas fueron creadas en MongoDB por lo cual no tienen una relación directa entre ellas a no ser SQL.

Candidatos

En esta tabla aguardamos la información de los candidatos, la estructura es:

- id
- nombre (Nombre del candidato)
- url_facebook (Pagina de Facebook)
- url_twitter (Pagina de Twitter)

Facebook

En esta tabla se guardan los datos que se recolectaron de Facebook con web scraping, de cada candidato en la fecha cuando se consultaron.

- Likes (Numero de likes)
- Seguidores (Número de seguidores)
- fecha-hora (Fecha de cuando se consultó)
- id_candidato (Id del candidato que pertenece)

Twitter

En esta tabla se guardan los datos que se recolectaron de Twitter con su API, de cada candidato en la fecha cuando se consultaron.

- Seguidores (Número de seguidores)
- Favoritos (Numero de Favoritos)
- fecha-hora (Fecha de cuando se consultó)
- id_candidato (Id del candidato que pertenece)

Puntajes_totales

En esta tabla se guardaron todos los puntajes totales que son la suma de los likes, seguidores de Facebook y twitter y favoritos, en la fecha cuando se consultaron.

- Puntaje (Puntaje total)
- fecha-hora (Fecha de cuando se consultó)
- id_candidato (Id del candidato que pertenece)

Todos los códigos empleados.

Servidor

Main.py

En esta parte se implementan las librerías necearías para el web Services y que funcione la conexión con mongo DB y las funciones que necesitamos de otros archivos, la configuración para configurar los CORS y el método para crear la BD y arrancar el servidor en el puerto 8080.

```
🍦 main.py 🛛 🗙
e main.py
        import bottle
        import pymongo
        import random
      from bottle import route, run, template
from bottle import response
 14 from pymongo import MongoClient
        from candidatosTwitter import apiTwitter
 18 \qquad {\sf from\ candidatosFacebook\ import\ webScrapingFacebook}
 from mongoDB import crearBD
from mongoDB import almacenarAPIWS, almacenarPuntajesTotales, getCandidatos
       from mongoDB import getPuntajesAcumalos, getTweet, getPuntosTotales
      #Variables Globales
cliente = MongoClient()
      db = cliente['BAD-Motor']
       app = bottle.app()
       def enable_cors(fn):
          def _enable_cors(*args, **kwargs):
               response.headers['Access-Control-Allow-Origin'] = '*'
response.headers['Access-Control-Allow-Methods'] = 'GET, POST, PUT, OPTIONS'
response.headers['Access-Control-Allow-Headers'] = 'Origin, Accept, Content-Type, X-Requested-With, X-CSRF-Token'
               if bottle.request.method != 'OPTIONS':
                      return fn(*args, **kwargs)
          return _enable_cors
        def main():
          #Crear la base de datos de los candidatos
           crearBD()
#Arranque del Web Services
             app.run(port=8080)
                                              --- Web Services
```

Ruta principal del Web Services que muestra la información, de las rutas en este.

Esta es la función para la ruta que muestra información que existe en la base de datos y se crea el json que regresara.

```
@route('/consulta')
     @enable_cors
53 v def consulta info():
         response.headers['Content-type'] = 'application/json'
         #JSON a enviar
         datos = {
              '1': {
                  'likes_F': 0,
                  'seguidores_F': 0,
                  'seguidores_T': 0,
                  'favoritos_T': 0,
                  'twett': '',
                  'puntajes': []
                  'likes_F': 0,
                  'seguidores_F': 0,
                  'seguidores_T': 0,
                  'favoritos_T': 0,
70
                  'twett': '',
                  'puntajes': []
74 🗸
                  'likes_F': 0,
                  'seguidores_F': 0,
76
                  'seguidores_T': 0,
                  'favoritos_T': 0,
                  'twett': '',
79
                  'puntajes': []
              },
82
```

En esta parte extraemos el número de likes, seguidores y favoritos de twitter y de Facebook, los sumamos para obtener como va su progreso de cada candidato.

Sigue sacando todos los puntajes totales de la base de datos y con estos puntos se hace una regresión lineal, para predecir cómo serán los resultados en 7 días más y regresamos el json al cliente.

```
#Analizando puntajes totales
          puntosTotales = getPuntosTotales()
          for i in range(3):
              i += 1
             #Variables regresion lineal
              n = len(puntosTotales[str(i)])
              xy = 0
             x = 0
              y = 0
              x2 = 0
118 🗸
              for punto in puntosTotales[str(i)]:
                 xy += int(punto['orden']) * int(punto['puntaje'])
                 x += int(punto['orden'])
                  y += int(punto['puntaje'])
                  x2 += int(punto['orden']) * int(punto['orden'])
             #Formulas de regresion lineal
              a = ((n*xy) - (x*y)) / ((n*x2) - (x*x))
              b = (y - (a*x)) / n
              for j in range(7):
                  j += 1
                  y = int((a*(n+j)) + b)
                  puntosTotales[str(i)].append({
                      'puntaje': y,
                      'fecha-hora': '+'+str(j)+' dia',
                      'orden': n+j
          datos['1']['puntajes'] = puntosTotales['1']
          datos['2']['puntajes'] = puntosTotales['2']
          datos['3']['puntajes'] = puntosTotales['3']
          return datos
```

En esta función es la ruta cuando recolectamos datos de internet, con la API de Twitter y haciendo Web Scraping de páginas de Facebook, los almacenamos en la base de datos, sumamos los puntajes obtenidos en las paginas y los guardamos en los puntajes totales, obtenemos la información de los candidatos y lo juntamos con la información recolectada para regresarla al cliente en json.

```
149 @route('/recoleccion/api-webScraping')
150 @enable_cors
151 v def data mine():
response.headers['Content-type'] = 'application/json'
     #Peticiones a API y Web Scraping
datosTwitter = apiTwitter()
       datosFacebook = webScrapingFacebook()
       datos = almacenarAPIWS(datosTwitter, datosFacebook)
       totales = [0 ,0 ,0]
for i in range(3):
           totales[i] = int(datos['api'][i]['seguidores']) + int(datos['api'][i]['favoritos']) + int(datos['webScraping'][i]['likes']) + int(datos['webScraping'][i]['seguidores'])
        datos = almacenarPuntajesTotales(totales, datos)
        datos['candidatos'] = getCandidatos()
171 v for ele in datos['api']:
           ele['_id'] =
        for ele in datos['webScraping']:
           ele['_id'] =
175 v for ele in datos['puntajesTotales']:
          ele['_id'] = '
        return datos
181 v if __name__=="__main__":
182 main()
```

MongoDB.py

En este archivo hacemos la conexión con la base de datos y guardamos o consultamos la información almacenada.

En esta parte importamos la librería para la conexión, y con la función creamos la base de datos, con la información de los candidatos.

En esta función almacenamos los datos de la API y del Web Scraping en las tablas de Facebook y Twitter.

Almacenamos los puntajes totales en la base de datos en la tabla con el mismo nombre.

Obtenemos los datos del candidato.

```
70 v def getCandidatos():
71 print('Conectando y Almacenando con Mongo DB . . .')
72 coleccion = db['candidatos']
73 #Optener a los candidatos
74 candidatos = []
75 v for candidato in coleccion.find({}, {'_id': 0}):
76 candidatos.append(candidato)
77 return candidatos
```

Obtenemos los puntajes acumulado de la tabla de Facebook y twitter.

Obtenemos los puntajes totales de cada candidato y ponemos un índice llamado orden para hacer la regresión lineal.

```
121 v def getPuntosTotales():
         print('Conectando y Almacenando con Mongo DB . . .')
          coleccion = db['puntajes_totales']
          datos = {
             '1': [],
             '2': [],
              '3': [],
          print('Extrayendo datos almacenados . . .')
          for i in (range(3)):
              i += 1
             orden = 1
             for puntos in coleccion.find({'id_candidato': i}, {'_id': 0, 'id_candidato': 0}):
                  puntos['orden'] = orden
                  datos[str(i)].append(puntos)
                  orden += 1
          return datos
```

CandidatosFacebook.py

Explicando la parte para obtener información de la red social de Facebook, primero se deben importar dos librerias que es la de request y beautifulsoup4, ya que la API que nos brinda fb se encuentra un poco limitada al acceso, por lo cual se concluyo en hacer WebScrapping a Facebook.

Esta parte será mandado a llamar desde otro código rutinariamente, ya que la función de este es regresar un array de los datos obtenidos de fb.

Iniciando a describir el código se tienen definidos las url's de los 3 candidatos que se tienen hasta el momento, donde obtendrá los likes y los seguidores basándonos en la estructura que tiene Facebook para mostrar la información, una vez tomada dichos datos se retornarán al código principal, donde se insertarán en la base de datos

```
# pip install beautifulsoup4
import requests
from bs4 import BeautifulSoup
def webScrapingFacebook():
   print('Web Scraping de paguinas estaticas de facebook . . .')
   urls.append("https://www.facebook.com/pg/raulmoronorozco/community/?ref=page_internal")
    urls.append("https://www.facebook.com/pg/CarlosHerreraSi/community/?ref=page_internal")
     urls.append("https://www.facebook.com/pg/juanAntonioMdelaMora/community/?ref=page_internal")
    candidatos=[]
    cont=0
     for url in urls:
        datos=[]
        cont=cont+1
        datos.append(str(cont))
         print('\tConectando con la paguina: '+url+' . . .')
             pag = requests.get(url)
            html = BeautifulSoup(pag.content, 'html.parser')
            conexion = True
           print('\tNo puede conectarse a la página '+url)
            conexion = False
         if conexion:
             print('\tAnalizando paguina: '+url+' . . .')
             comunidad = html.find('div', class_='clearfix _ikh _3xol')
```

```
#Obtener los likes de la página
seccion = (comunidad.find_all('div', class_='_3xom'))
seccion_likes=seccion[0].text
likes=seccion_likes.replace("\xa0mil", "000")

seccion_seguidores=seccion[1].text
seguidores=seccion_seguidores.replace("\xa0mil", "000")

likes=likes.replace(".", "")
seguidores=seguidores.replace(".", "")

datos.append(likes)
datos.append(seguidores)

candidatos.append(datos)

else:
datos.append("")
datos.append("")
candidatos.append(datos)

return candidatos

return candidatos
```

CandidatosTwitter.py

El objetivo dentro de este programa Python es consumir la api de Twitter para poder obtener información específica de cada candidato a la gobernatura de Michoacán, para poder realizar esto hacemos uso de la librería tweepy, lo primero que hacemos es autenticarnos con nuestros 4 tokens que se muestran de la línea 8 a 11. Una vez hecho lo anterior obtenemos toda la información de cada candidato que nos proporciona la api, después por cada candidato declaramos un arreglo e insertamos para cada uno almacenamos sus followers y favoritos recibidos, una vez hecho esto retornamos un arreglo que contiene los 3 arreglos de los candidatos.

```
import tweepy
import json
def apiTwitter():
    print('Accediendo a la api de Twitter . . .')
    consumer key="oU3eRnB@ENoGq124quINXdrAZ"
    consumer secret="2QMnOEuGbG6EUXccaOndksaTvT4LAmrKJ838bb2IuCZASonxCg"
    access_token = "418197577-pXUdPpPC6NJDbmaAgQSpNRkvEMSt1xep5foYjBqk"
    access token secret = "JxArOpeZg8B29dRsdnRGSQYWOtC1xeWe1DbrZULFNA6s0"
    auth = tweepy.OAuthHandler(consumer key, consumer secret)
    auth.set_access_token(access_token, access_token_secret)
    api = tweepy.API(auth, wait on rate limit=True,
    wait on rate limit notify=True)
    data = api.get_user("CarlosHerreraSi")
    dataRM = api.get user("raulmoron0")
    dataJM = api.get_user("Magana_DeLaMora")
    Candidatos = []
    #Carlos Herrera
    print('\tCarlos Herrera . . .')
   CarlosHerrera = []
    followers = str(data.followers_count)
    noFavoritos = str(data.favourites_count)
    uTweet = str(data.status.text)
   CarlosHerrera.append("2")
    CarlosHerrera.append(followers)
    CarlosHerrera.append(noFavoritos)
   CarlosHerrera.append(uTweet)
```

```
#Raúl Morón Orozco
         print('\tRaúl Morón Orozco . . .')
         RaulMoron = []
         RMfollowers = str(dataRM.followers count)
         RMnoFavoritos = str(dataRM.favourites count)
         RMuTweet = str(dataRM.status.text)
41
         RaulMoron.append("1")
42
         RaulMoron.append(RMfollowers)
         RaulMoron.append(RMnoFavoritos)
44
         RaulMoron.append(RMuTweet)
47
         #Juan Antonio Magaña de la Mora
         print('\tJuan Antonio Magaña de la Mora . . .')
         JuanMagania = []
         JMfollowers = str(dataJM.followers count)
         JMnoFavoritos = str(dataJM.favourites count)
         JMuTweet = str(dataJM.status.text)
         JuanMagania.append("3")
         JuanMagania.append(JMfollowers)
         JuanMagania.append(JMnoFavoritos)
         JuanMagania.append(JMuTweet)
         print('\tAnalizando datos . . .')
         Candidatos.append(RaulMoron)
         Candidatos.append(CarlosHerrera)
         Candidatos.append(JuanMagania)
         return Candidatos
```

Cliente

Documentos HTML, para mostrar los datos recolectados, que contiene a los candidatos y las graficas de progreso y comparación de puntajes.

Index.html

```
| cliv class="row graphics" |
| ch2 class="titulod">
| ch2 class="titulod">
| cliv class="clind-12 col-1g-6 col-sm-12" id="columnchart_values" ></div)
| cliv class="clind-12 col-1g-6 col-sm-12" id="columnchart_values" ></div)
| cliv class="titulod">
| cliv class="clind-12 col-1g-6 col-sm-12" id="columnchart_values" ></div)
| cliv clind-12 col-1g-6 col-sm-12" id="columnchart_values" ></div)
| clind-12 col-1g-6 col-sm-12" id="col-sm-12" id="columnchart_values" ></div)
| clind-12 col-1g-6 col-sm-12" id="col-sm-12" id="col-sm-12"
```

Principal.js

En esta función hacemos cuando el documento se haya cargado completamente hacemos una petición AJAX, al servidor la cual recibe el json con la información, la pone en su respectivo lugar dentro de index.html.

```
$(document).ready(function(){
   $.ajax({
      type: "get",
      url: "http://127.0.0.1:8080/consulta",
      data: {},
      dataType: "json",
      success: function (json) {
         console.log(json);
         //Carlos Herrera Tello
         let templateCH =
         Followers en Twitter: `+ json["1"].seguidores_T +`
         Favoritos en Twitter: `+ json["1"].favoritos_T+`
         Total de likes en Fb: `+ json["1"].likes_F +`
         Total de seguidores en Fb: `+ json["1"].seguidores_F +`
         $("#datosCH").html(templateCH);
         //Raúl Morón
         let templateRM = `
         Followers en Twitter: `+ json["2"].seguidores_T +`
         Favoritos en Twitter: `+ json["2"].favoritos_T+`
Total de likes en Fb: `+ json["2"].likes_F +`
         Total de seguidores en Fb: `+ json["2"].seguidores_F +`
         $("#datosRM").html(templateRM);
         let templateJM = `
         Followers en Twitter: `+ json["3"].seguidores_T +`
         Favoritos en Twitter: `+ json["3"].favoritos_T+`
         Total de likes en Fb: `+ json["3"].likes_F +`
         Total de seguidores en Fb: `+ json["3"].seguidores_F +`
         $("#datosJM").html(templateJM);
```

Llena la gráfica de puntos con los puntajes totales y la predicción a 7 días.

```
//Graphics
google.charts.load('current', {'packages':['corechart']});
google.charts.setOnLoadCallback(drawChart);
function drawChart() {
   let templateGraphics = ``;
    let graficaPuntajes = [
         ['Día', 'Carlos Herrera', 'Raúl Morón', 'Juan Magaña'],
    let longitud = json["1"]["puntajes"].length;
    for (let i = 0; i < longitud; i++) {
              json["1"]["puntajes"][i]['fecha-hora'],
             json["1"]["puntajes"][i]['puntaje'],
json["2"]["puntajes"][i]['puntaje'],
json["3"]["puntajes"][i]['puntaje']
         graficaPuntajes.push(col);
    var data = google.visualization.arrayToDataTable(graficaPuntajes);
    var options = {
        title: 'Presencia en redes sociales',
         curveType: 'function',
         legend: { position: 'bottom' },
         series: {
             0: { color: '#FFC300' },
1: { color: '#C70039' },
2: { color: '#0eb62f' },
    var chart = new google.visualization.LineChart(document.getElementById('curve_chart'));
    chart.draw(data, options);
```

Llena la gráfica de barras que muestra la comparación entre los puntajes de los candidatos.

```
/* Barnas */
google.charts.setOnLoadCallback(drawBarnas);
function drawBarnas() {
    var data = google.visualization.arrayToDataTable([
        ["Element", "Osnof"2"].seguidores_I + json["1"].likes_F + json["1"].seguidores_F), "#FFC300"],
        ["Radi Nonón", (json["2"].seguidores_I + json["2"].favoritos_I + json["2"].seguidores_F), "#C70930 "],
        ["Juan Magaña", (json["3"].seguidores_I + json["3"].favoritos_I + json["3"].likes_F + json["3"].seguidores_F), "#C70930 "],
        ["Juan Magaña", (json["3"].seguidores_I + json["3"].favoritos_I + json["3"].likes_F + json["3"].seguidores_F), "#C70930 "],
        ["Juan Magaña", (json["3"].seguidores_I + json["3"].likes_F + json["3"].seguidores_F), "#C70930 "],
        ["Sun Magaña", (json["3"].seguidores_I + json["3"].likes_F + json["3"].seguidores_F), "#C70930 "],
        ["Sun Magaña", (json["3"].seguidores_I + json["3"].likes_F + json["3"].seguidores_F), "#C70930 "],
        ["Sun Magaña", (json["3"].seguidores_I + json["3"].likes_F + json["3"].seguidores_F), "#C70930 "],
        ["Sun Magaña", (json["3"].seguidores_I + json["3"].likes_F + json["3"].seguidores_F), "#C70930 "],
        ["Sun Magaña", (json["3"].seguidores_I + json["3"].likes_F + json["3"].seguidores_F), "#C70930 "],
        ["Sun Magaña", (json["3"].seguidores_I + json["3"].likes_F + json["3"].seguidores_F), "#C70930 "],
        ["Sun Magaña", (json["3"].seguidores_I + json["3"].likes_F + json["3"].seguidores_F), "#C70930 "],
        ["Sun Magaña", (json["3"].seguidores_I + json["3"].likes_F + json["3"].seguidores_F), "#C70930 "],
        ["Sun Magaña", (json["3"].seguidores_I + json["3"].likes_F + json["3"].likes_F + json["3"].seguidores_I + json["3"].likes_F + json["3"].seguidores_I + json["3"].likes_F + json["3"].likes_F + json["3"].seguidores_I + json["3"].likes_F + json["3"].likes_F + json["3"].likes_F + json["3"].seguidores_I + json["3"].likes_F + json["3"].likes_F
```

Función clic del botón para consultar nuevos datos del internet a través de la API y el Web Scraping de Facebook, para esto hacemos una petición AJAX, al servidor para hacer esta acción y nos regrese una respuesta, para mostrar un mensaje.

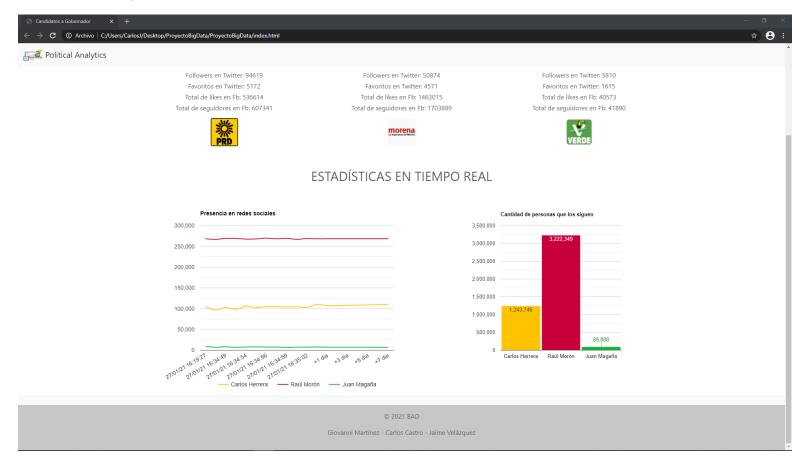
Capturas de pantalla de la impresión de resultados.

Los resultados que son mostraran son los que recabamos en una semana de extraer datos del internet.

En la primera parte podemos ver a cada candidato, con su información y numero de seguidores de facebook y de twitter, así como el numero de likes de Facebook y de favoritos de Twitter.



Por último, vemos las gráficas, la primera es del avance en redes sociales que a tenido un candidato, y posteriormente la predicción a futuro de como estará en redes sociales, seguidamente una grafica que compara la presencia de estos en redes sociales.



Como podemos ver en la grafica de comparaciones, el posible futuro ganador será **Raúl morón.** Por superar a los demás en presencia en redes sociales.