



# PROYECTO 1: ANALITICA DE TEXTO

Etapas 2

santiago rodriguez cruz y sebastián umaña peinado  
202013371 y 202013778

## Contenido

1. Proceso de automatización del proceso de preparación de datos, construcción del modelo, persistencia del modelo y acceso por medio de API.....	1
2.Desarrollo de la aplicación y justificación. Descripción del usuario/rol de la organización que va a utilizar la aplicación, la conexión entre esa aplicación y el proceso de negocio que va a apoyar (si aplica), y la importancia que tiene para ese rol la existencia de esta aplicación. ....	4
3.Trabajo en Equipo .....	7

## 1. Proceso de automatización del proceso de preparación de datos, construcción del modelo, persistencia del modelo y acceso por medio de API

En base a la etapa 1 se eligió el modelo mas certero, que corresponde a Support Vector Machine(Esto corresponde a la etapa 1)

```
svm_pipeline = Pipeline([
    ('tfidf', TfidfVectorizer()),
    ('svmclf', SVC(kernel='rbf'))
])

svm_pipeline.fit(X_train1, y_train1)
svmpred=svm_pipeline.predict(X_test1)
```

Ya para la etapa 2, exportamos el modelo.

```
#exportamos el modelo de SVM en joblib
dump(svm_pipeline, 'modelo_svm.joblib')
```

Utilizamos el framework Flask para el desarrollo de las APIs y de la página web. Por temas de sencillez y congruencia con la etapa 1, se hizo el preprocesamiento de datos desde la aplicación para que entre al pipeline del modelo como tal posteriormente.

```
# Función para procesar los datos
def process_data(text):
    return text.apply(porter_stemmer_spanish).apply(join_words)

def limpiar_texto(texto):
    tokenizado = word_tokenize(texto)
    procesado = procesar_texto(tokenizado)
    stemmizado = porter_stemmer_spanish(procesado)
    completado = join_words(stemmizado)
    return completado
```

Después de cargar el modelo que se exportó. Definimos las siguientes 3 plantillas para cargar las diferentes páginas que soporta la aplicación web

```

# Ruta principal para cargar la página HTML
@app.route('/')
def index():
    return render_template('index.html')

@app.route('/csv')
def csv():
    # Lógica para renderizar la plantilla HTML del formulario de carga de CSV
    return render_template('csv.html')

@app.route('/review')
def review():
    # Lógica para renderizar la plantilla HTML del formulario de carga de texto
    return render_template('review.html')

```

Se definieron 2 APIs. Una para predecir una entrada de texto individual/reseña y otra para cargar entradas de texto en masa por formato CSV. Las dos tienen métodos POST y GET.

El primer api tiene una función llamada `predecir_text`. En el primer bloque `if`, se verifica si la solicitud es del tipo POST. Esto significa que el cliente está enviando datos al servidor, como el contenido de un formulario. Si es así, se espera que los datos se encuentren en el cuerpo de la solicitud (`request body`) y se accede a ellos utilizando `request.json`, que interpreta el contenido como JSON y lo convierte en un diccionario de Python. Luego, se extrae el texto ingresado por el usuario utilizando la clave `'text'`.

En el bloque `elif`, se verifica si la solicitud es del tipo GET. Esto significa que el cliente está solicitando información del servidor, como una página web o datos específicos. Si es así, se espera que los datos estén en la URL de la solicitud y se pueden acceder a través de los argumentos de la URL utilizando `request.args.get('text', '')`. Esto obtendrá el valor del parámetro llamado `'text'` en la URL, o un valor predeterminado vacío si no se proporciona ese parámetro. Ya hechas las dos solicitudes, se limpia el texto y se envía al modelo para predecir.

```

# Rutas para la predicción de texto
@app.route('/predict_text', methods=['POST', 'GET'])
def predict_text():
    if request.method == 'POST':
        data = request.json
        text = data['text']
    elif request.method == 'GET':
        text = request.args.get('text', '')

    text = tr.limpiar_texto(text)
    prediction = svm_model.predict([text])[0]
    prediction = int(prediction) # Convertir la predicción a tipo int
    return jsonify({'prediction': prediction})

```

Para el lado de la función `predict_csv` tiene un funcionamiento muy similar a la función de predecir reseñas individuales, la diferencia principal radica en que esta función tiene incluida una limpieza básica del CSV antes de aplicar el preprocesamiento definido en `tratado.py`. En las dos funciones se usa la función `jsonify` con el objetivo de que lo retornado se devuelva como un json y se pueda manejar con el javascript de la lógica de la aplicación.

2.Desarrollo de la aplicación y justificación. Descripción del usuario/rol de la organización que va a utilizar la aplicación, la conexión entre esa aplicación y el proceso de negocio que va a apoyar (si aplica), y la importancia que tiene para ese rol la existencia de esta aplicación.


Dentro de la organización de COTELCO la aplicación está destinada principalmente a las subdivisiones/Capítulos de COTELCO y en menor medida a los hoteles de esa área en específico. El proceso al que podría estar relacionado son las investigaciones sectoriales que hace COTELCO a través del CPTUR, que consisten en análisis y seguimiento del turismo en ciertos sectores. La aplicación

valdría para soportar los reportes de las investigaciones de factores agravantes o beneficiosos en las áreas hoteleras con análisis predictivo.




Esta aplicación es importante para el CPTUR pues les permite validar la información que ellos encuentren al momento de hacer los procesos asociados a sus investigaciones de sector. Por ejemplo, si COTELCO y el CPTUR deciden hacer entrevistas a habitantes de cierta zona preguntándoles su opinión de la seguridad del área, se puede validar dichas entrevistas por medio de reseñas ,sean fabricadas o ya existentes, que contengan lo que se socializó en dichas entrevistas y corroborar resultados.

Ya definido el actor de la organización con los estudiantes de estadística, orientamos el diseño estético de la aplicación hacia COTELCO. Las funcionalidades de la aplicación se limitaron a predecir reseñas individuales y en masa por medio de carga de archivos. Se decidió implementar solo estas funciones debido a que es lo más pertinente en términos de como se entreno el modelo y del entendimiento del negocio. Aunque se pensó en la predicción de reseñas incluir los sectores asociados a esa reseña, no se implementó debido a la falta de datos para implementar esto de manera apropiada por parte del negocio.


Este es el resultado final de la aplicación:

[Inicio](#) [Reseña](#) [Cargar Reseñas](#)

**Bienvenidos a la plataforma de evaluación de calidad hotelera - PECH**



Esta plataforma tiene el propósito de evaluar los resultados obtenidos de su investigación sectorial en la nación colombiana por medio del análisis predictivo de reseñas individuales y en masa.


[Inicio](#) [Reseña](#) [Cargar Reseñas](#)

**Describe la reseña a la cual quiere predecir su calificación**

Ingresa el texto para predecir

Predecir Texto

ID	Reseña	Calificación
----	--------	--------------

<div>  <a href="#">Inicio</a> <a href="#">Reseña</a> <a href="#">Cargar Reseñas</a> </div>		
<b>Cargue el archivo de reseñas con formato CSV</b>		
<div> <a href="#">Seleccionar archivo</a>   <a href="#">partición csv</a> </div> <div> <a href="#">Cargar CSV</a> </div>		
ID	Reseña	Calificación
1	1 a primera noche nos encontramos en la habitación con un rojo (5) de cucarachas muertas y la atmósfera muy sucia... que por cierto nunca limpiaron hasta que reclamamos... La cena de 10 de año fue un bulte que acabó en un tremendo desorden por el excesivo número de mesas vacías, no se podía caminar para ir a servirse la comida, para luego encontrarte con bandejas vacías, mi esposa y yo y otra pareja "NO CENAMOS" porque nos encontramos con puras bandejas vacías... pagamos \$2000 pesos por por panes por nada!! El servicio del restaurante pésimo, pésimo,pésimo...se tardaron muchísimo tiempo en atendernos, nos trajeron el desayuno en platos, nunca nos dieron el café que pedimos desde que llegamos... un pésimo servicio!! El lugar vacío y después de reclamar lo sustituyeron por...afórra...	1
2	A una calle de catedral con platillos tradicionales, tipo Gourmet, de buen sabor y calidad. Si bien ofrecen una carta con los platillos más representativos de cada temporada como mole de caderas, potolito, chinchales, chiles en nogada, escamoles y cerezas, son principalmente para degustación, no...Más	5
3	Porcosos muy resacaibles-Agua con sabor a omei Muy delicias los mocos. No losa los respectivos Hay regentes opositos	3
4	Cartagena encanta. De todas las ciudades coloniales que hemos visitado es la más linda y mejor conservada. Recomendamos recomendar tanto de día como de noche ya que son visiones distintas que vale la pena vivir. Llena de vida, tiendas, restaurantes, colores, vibrante, calles y rincones que en cada momento guardan una sorpresa. Mi calificación habría sido excelente sino fuera por algunos detalles que deberían mejorar: Lamentablemente falta preocupación por la limpieza de las calles (problema que se repite en el resto de la ciudad) y un mejoramiento en los alcantarillados de las aguas servidas. Además por la estrechez de las calles deberían habilitar vías exclusivas para peatones.	5
5	Buenos con mucha fealdad de disfrutar el espectáculo de luz y sonido pero la verdad nos decepcionó, lo abusan en una espina a la entrada de los cultos y no al centro por lo que no pueden apreciar casi nada las personas se paran a tomar fotografías debido a que no se alcanza a ver nada o están sentados al inicio y al final del espectáculo no dejan prendidas las luces para que puedas tomar fotografías, al contrario queda todo en oscuridad y las personas de la tercera edad tiene problemas para salir pues no luminan en el camino para salir, deberían poner algunas sea unas antorchas para iluminar la verdad no lo recomiendo es una pena que una maravilla como lo es chichen itza tenga un espectáculo tan carente	4
6	El sábado 12 de 11 de 2016, desde que llegue al personal tiene el peor trato, no saben inglés, y al parecer tampoco español, 5 veces me preguntaron si mi reservación era así, como si después de la 3 o 4 vez que me preguntaron mi respuesta sería diferente, por más que les compré conoucher, itinerario y pago, no sabían en donde estaba mi reservación, y decían que era hasta dentro de un mes, chová fue error de ellos, pero nunca lo admitieron. Llegue a las 12 pm y no me encontraron habitación hasta la 1pm, prometieron llevalla antes y no la traen hasta después de las 3 pm debido a la ocupación que tenían, el hotel y los precios son de lujo pero el trato es de lo peor fue que habitar con 3 incomprensivos y nadie podía hacer nada al gerente nunca llegó, el supervisor nunca apareció, y simplemente volvíamos para otro lado cada vez que les preguntaba por mi habitación, practicamente 4 horas perdidas de mi día, que no se pudo recuperar en los restaurantes los meses, subiendo la mañana volvíamos para otro lado al momento de pedir servicio, por mas desorganizado que estaba el lugar no ofrecían ni una mesa, pero si ven a un estado ordenado estar hasta ahorrara roja la leyenda, no entiendo por que en este país en esta zona en especial hay tan mal trato a sus propios pasajeros, totalmente desagradable mi experiencia en este hotel, nada recomendable si eres de mexico. Y NO HAY PLAYA COMO LO DICE EN INTERNET. TE TIENEN QUE TRASLADAR A EL OTRO HOTEL PARA FESO	2

### 3.Trabajo en Equipo

Se asignaron los siguientes roles:

Líder de Proyecto:Santiago Rodriguez Cruz

Ingeniero de software responsable de desarrollar la aplicación  
final:Sebastián Umaña Peinado

Ingeniero de datos:Sebastián Umaña Peinado

Ingeniero de software responsable del diseño de la aplicación y  
resultados:Santiago Rodriguez Cruz, Sebastián Umaña Peinado



### Cronograma Reuniones:

Reunión de Lanzamiento y Planeación: 8/04/2024 duración 20m

Reunión de Seguimiento: 12/04/2024 duración 1h

Reunión De Finalización: 19/04/2024 duración 1h30m

### Tareas Realizadas:

El desarrollo de cada tarea se hizo en conjunto entre los dos. La última tarea se hizo con el apoyo del grupo de estadística.

Exportar Modelo y realizar funciones externas de tratamiento de texto: 1h

Desarrollo de la API: 1h

Desarrollo de requerimientos de la aplicación web: 3h