Texto

Descripción generada automáticamente con confianza media

**Resúmenes de desafío**

[Desafío 1 (Word2vect) 2](#_Toc118038248)

[Resultados 2](#_Toc118038249)

[Desafío 2 (Bot basado en reglas con DNN + Spacy) 4](#_Toc118038250)

[Resultados 4](#_Toc118038251)

**Especialización en Inteligencia Artificial**

**Materia** : Procesamiento de Lenguaje Natural

**Año** : 2022

**Docentes** :

Msc. Rodrigo Cardenas Szigety (rodrigo.cardenas.sz@gmail.com)

Esp. Ing. Hernán Contigiani hernan4790@gmail.com

**Alumno** : Marcelo Chichiri

# Desafío 1 (Word2vect)

Utilizando herramientas de numpy implementar la comparación de documentos por medio de TF-IDF.

Los documentos propuestos son 'que dia es hoy', 'martes el dia de hoy es martes', 'martes muchas gracias', pero el programa sirve para cualquier número de documentos.

## Resultados

Los resultados fueron satisfactorios. Para los tres documentos indicados son las siguientes:

Calendario

Descripción generada automáticamente con confianza media

El resultado final con la comparación entre documentos es el siguiente:

Tabla

Descripción generada automáticamente

# Desafío 2 (Bot basado en reglas con DNN + Spacy)

Utilizando las técnicas vistas en clase, y tomando el ejemplo de los bots utilizados, implementar un bot para un uso particular.

## Resultados

Para este desafío usé un dataset para un bot de un servicio de venta y reclamos de un servicio de TV por cable (Cablin).

Se definieron 11 clases ('Subscripcion', 'agradecimientos', 'baja', 'consulta\_pago', 'contacto', 'despedida', 'hablar\_humano', 'nombre', 'reclamos\_cobro', 'reclamos\_servicios', 'saludos') con un vocabulario de 93 términos.

Para implementar este desafío se realizaron las

* Se eliminaron mas signos de puntuación, como "~@#$%^&\*()\_+<>"
* Para ajustar un poco el modelo tuve que agregar algunos patterns que no había considerado inicialmente. Esto hizo que el modelo mejorase su respuesta.
* Se probó sin stop words y con stop words. El resultado con stop words empeoró para algunos casos, ya que se eliminaban algunos terminos como 'no', 'como' y 'este' que hacían que algunos documentos no tuviera sentido. Para evitar este problema se eliminó estos tres términos de la lista de stop words.

El resultado, dentro de este modelo básico, creo que resultó muy satisfactorio:

Hola

Q: Hola

BOT: Hola!, Soy Cablin ¿En te puedo ayudar?

Buen día

Q: Buen día

BOT: Hola!, Soy Cablin ¿En te puedo ayudar?

Como estas?

Q: Como estas?

BOT: Hola!, Soy Cablin ¿En te puedo ayudar?

tengo problema con el servicio

Q: tengo problema con el servicio

BOT: Por problemas técnicos llama al 011-000-1111 (opcion 1) para agendar una visita del técnico

el cable se corta

Q: el cable se corta

BOT: Por problemas técnicos comunicarte al 011-000-1111 (opcion 1) para programar una vista del técnico

y me cobraron de mas

Q: y me cobraron de mas

BOT: Para consultar por facturación podés llamar al 011-000-1111 (opción 4) para hablar con un representante

tampoco me llegó la factura

Q: tampoco me llegó la factura

BOT: Para consultar por facturación podés podés comunicate al 011-000-1111 (opción 4) para hablar con un representante

quiero saber cuales son los planes

Q: quiero saber cuales son los planes

BOT: Mi nombre es Cablin

como hago si me quiero dar de baja?

Q: como hago si me quiero dar de baja?

BOT: Para darte de baja o cambiar tu subscripción llamá al 011-000-1111 (opción 2) y te atenderá un representante

Ahora estoy pagando por debito, que otros medios de pago hay?

Q: Ahora estoy pagando por debito, que otros medios de pago hay?

BOT: Visita nuestra web www.cablin.com.ar para ver todas las formas de pago

quien sos?

Q: quien sos?

BOT: Mi nombre es Cablin

Puedo hablar con una persona?

Q: Puedo hablar con una persona?

BOT: Para hablar con un representante podés llamar al 011-000-1111

Con quien estoy hablando?

Q: Con quien estoy hablando?

BOT: Estas hablando con Cablin

gracias!!

Q: gracias!!

BOT: Gracias a vos por comunicarte con Cablin

adios

Q: adios

BOT: Hasta luego!, gracias por tu visita

# Desafío 3 (Custom embedddings con Gensim)

El objetivo es utilizar documentos / corpus para crear embeddings de palabras basado en ese contexto. Se utilizará canciones de bandas para generar los embeddings, es decir, que los vectores tendrán la forma en función de como esa banda haya utilizado las palabras en sus canciones.

## Resultados

Para este desafío se usó un corpus de aproximadamente 40 canciones de Joan Manuel Serrat. El dataset fue creado a mano. Las letras fueron obtenidas de <https://www.musica.com/letras.asp?letras=7599>

* La temática de los temas de cantautor es muy variada por lo cual los términos aparecen en contextos variados.
* Se estableció en 2 la cantidad de repeticiones mínimas del término para ser tomada en cuenta. Esto es principalmente porque hay términos que se repiten poco pero dan significado a otras que se repiten mas.
* En el gráfico de embedding no se puede ver la misma relación, lo cual puede deberse a que fue necesario usar solo dos dimensiones para poder presentarlo.
* Se prueba el modelo con varios términos como **amigos**, **camino**, **flores**, **piratas** y **hombres**
* Se comprueba que los términos se pueden relacionar con distintos párrafos de canciones de cantautor.