

Aplicación web que clasifica opiniones para determinar la postura sobre temas políticos/sociales de los usuarios de la red social Twitter en México.

Trabajo Terminal No. _____

Alumnos: *Carpintero Mendoza Marcos Mauricio, Reyes Reyes Karla Alexia

Directores: Kolesnikova Olga

e-mail: marcos_c79@hotmail.com

Resumen - El presente documento propone una aplicación web que pretende dar a conocer la postura general de los usuarios de Twitter, la red social más popular de microblogging, sobre temas de interés político y de impacto social con la finalidad de ofrecer una métrica de la opinión pública haciendo uso de representaciones gráficas. Se utilizarán técnicas de análisis de sentimientos basadas en Procesamiento de Lenguaje Natural y algoritmos de Aprendizaje de Máquina.

Palabras clave - Aplicación web, Aprendizaje de Máquina, Procesamiento de Lenguaje Natural.

1. Introducción.

El lenguaje escrito siempre ha sido de gran importancia para la humanidad: es una forma básica de comunicación, es útil para registrar contenido relevante, nos permite plasmar nuestras ideas, esto por mencionar algunos de sus usos. Cabe mencionar que implícitamente en ciertos textos, sin importar lo corto o largo que sean, en estos se transmite información adicional tal como emociones u opiniones.

Dentro del Procesamiento de Lenguaje Natural, el análisis de sentimientos es el estudio de las opiniones, actitudes y emociones individuales de una persona usando técnicas computacionales para poder interpretar su postura ante temas específicos a partir del lenguaje humano, en este caso específicamente en representación de texto. A través de los años, las investigaciones dentro de esta ciencia han ido mejorando las técnicas y modelos para obtener mejores resultados: comenzando con un enfoque léxico y ahora empleando un enfoque de aprendizaje de máquina. [1]

Por otro lado, se sabe que con el paso de los años, las redes sociales han ido ganando popularidad e incrementado diariamente su número de usuarios, por lo que la gran cantidad de publicaciones que se realizan en éstas pueden ser de utilidad para las organizaciones, después de haber pasado por el procesamiento necesario. El uso de la redes sociales permite a la población expresarse de una forma más liberal, por lo que la opinión pública a través de este medio, con un correcto análisis de lenguaje natural, podría verse desde una perspectiva más real y lo más importante conservando el anonimato de los usuarios de la red social dentro de nuestra métrica de resultados. [2]

Finalmente, una aplicación enfocada a la clasificación de opiniones debe de resolver dos desafíos en general; primero, encontrar una solución óptima para el procesamiento de gran cantidad de datos para lograr los objetivos que se establezcan para el sector beneficiario, y segundo, ser capaz de identificar el sentimiento más adecuado para lo que se quiere transmitir en el texto analizado. [2]

Específicamente nuestra aplicación está enfocada en mostrar la opinión de los mexicanos en temas políticos-sociales, para que las organizaciones políticas tomen acciones que beneficien los problemas que expresen la población.

2. Objetivo.

Desarrollar una aplicación web utilizando técnicas de análisis de sentimientos junto con aprendizaje de máquina para poder clasificar las opiniones que tengan un enfoque político/social de la población mexicana a través de

twitter, sirviendo así de fuente de consulta a organizaciones políticas y/o sociales. Esto con el fin de que las organizaciones entiendan tanto el impacto que tiene su imagen y las acciones que toman para con la ciudadanía, y finalmente atiendan las necesidades de los ciudadanos.

Objetivos específicos:

- Conseguir un porcentaje mayor a 75% de certeza.
- Crear un catálogo de consulta para temas basados en tweets cuya publicación no sea mayor a dos semanas con respecto al día de consulta.
- Lograr el manejo de una cantidad de entre 8,000 a 10,000 tweets en el conjunto de entrenamiento para el modelo de aprendizaje.

3. Justificación.

Actualmente la forma más común de recabar la opinión pública es mediante el uso de encuestas, las cuales a veces conllevan a sesgo debido a aspectos implícitos tales como [3]:

- Número limitado de encuestas a realizar: Se destina un número fijo de encuestas para aplicar en un lugar en específico dejando fuera a cierto número de la población.
- Método de aplicación presencial: Algunas encuestas son diseñadas para que se atiendan de forma presencial obligando la asistencia de la población.
- Número insuficiente de participantes: Las personas suelen ser renuentes a participar, lo que da como resultado que la información recabada no sea totalmente confiable para dar el resultado de interés.

Así mismo, hay ocasiones en las que no hay encuestas para ciertos temas políticos ignorando completamente la opinión pública, llegando a provocar disgusto en la población.

Nosotros proponemos una aplicación web que funcione como complemento, de consulta de la opinión pública, a las encuestas. Para esto utilizaremos datos de la red social Twitter, ya que consideramos que en esta red social los usuarios expresan libremente su opinión sobre gran variedad de temas. Además, es importante tomar en cuenta que dentro de los usuarios de la red social se encuentran: presidentes, diputados, medios de comunicación, etcétera.

4. Productos o Resultados esperados.

Al finalizar el trabajo terminal tendremos como resultado una aplicación web capaz de clasificar opiniones en por lo menos 3 categorías: positivas, negativas y neutrales, haciendo uso de una arquitectura cliente-servidor, junto a herramientas de inteligencia artificial para el procesamiento de los datos.

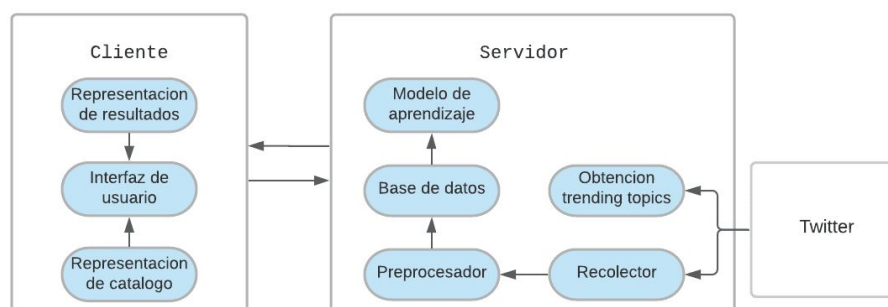


Figura 1. Arquitectura Cliente-Servidor

Probaremos diferentes algoritmos de aprendizaje de máquina, por ejemplo, Máquina de Soporte de Vectores [4], Multilayer Perceptron, Naive Bayes, entre otros para poder determinar cual es el que mejores resultados nos ofrece y así obtener la mejor precisión posible; de igual manera que el algoritmo nos permita clasificar las opiniones en las categorías antes mencionadas.

- Aplicación web.
- Código fuente
- Documentación del desarrollo.
- Manual de usuario.

5. Metodologías.

La metodología a emplear para la elaboración del proyecto será SCRUM dado que la filosofía que maneja nos ofrece las siguientes ventajas: [5]

- Permite que el desarrollo sea flexible, creativo y productivo.
- El trabajo se maneja por iteraciones/incrementos para lograr las metas que se establecen en cada sprint.
- Maneja una comunicación continua entre los miembros del equipo para trabajar fluidamente y así evitar errores que puedan retrasar las tareas planteadas.
- Implementa una pila del producto (Product backlog) la cual contiene las características, funcionalidades, requisitos, mejoras y correcciones del producto.

Nuestro principal reto es lograr un porcentaje de certeza lo más alto posible para conseguir la mejor clasificación de las opiniones que tengamos que analizar y consideramos que haciendo uso de esta metodología podemos ir mejorando el desempeño de nuestro algoritmo de aprendizaje a lo largo del desarrollo del proyecto, esto porque es un proceso iterativo, escalable y modular.

Existen diferentes reuniones en cada iteración donde el tema principal es el desarrollo del proyecto para poder identificar los errores, fallas, tareas cumplidas, nuevos requerimientos; por mencionar algunos aspectos del desarrollo, de esta forma se logra mejorar en equipo los puntos débiles que puedan afectar el desarrollo del proyecto y mantener en el mismo nivel las características en donde el equipo tiene un óptimo desempeño.

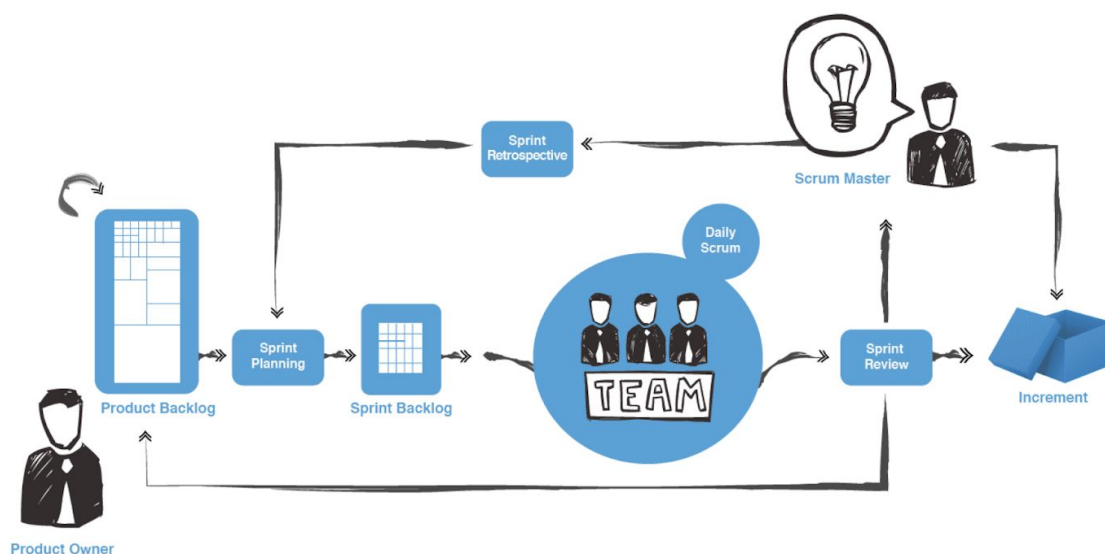


Figura 2. Ciclo de desarrollo SCRUM [6]

6. Cronograma

La duración promedio de cada sprint es de cuatro semanas, esto porque hay un grupo destinado para cada área de desarrollo y en nuestro caso tomaremos el papel de los distintos personajes necesarios dentro de la metodología de SCRUM. Es por esto que la duración de nuestro sprint tiene un rango de cuatro semanas a seis semanas, esto dependiendo del grado de dificultad de los requerimientos que se necesiten.

CRONOGRAMA Nombre del alumno: Carpintero Mendoza Marcos Mauricio

TT No:

Título del TT: Aplicación web que clasifica opiniones para determinar la postura sobre temas políticos/sociales de los usuarios de la red social Twitter en México.

[illegible]

CRONOGRAMA Nombre del alumno: Reyes Reyes Karla Alexia

TT No:

Título del TT: Aplicación web que clasifica opiniones para determinar la postura sobre temas políticos/sociales de los usuarios de la red social Twitter en México.

[illegible]

7. Referencias

- [1] Bhadane, Chetashri & Dalal, Hardi & Doshi, Heenal. (2015). Sentiment Analysis: Measuring Opinions. *Procedia Computer Science*. 45. 808-814. 10.1016/j.procs.2015.03.159.
- [2] Stieglitz, S., Dang-Xuan, L. Social media and political communication: a social media analytics framework. *Soc. Netw. Anal. Min.* 3, 1277–1291 (2013). <https://doi.org/10.1007/s13278-012-0079-3>
- [3] Khan, Sarwar & Khan, Muzammil & Ran, Qiong & Naseem, Rashid. (2018). CHALLENGES IN OPINION MINING, A COMPREHENSIVE REVIEW. 33. 123-135.
- [4] Bing Liu (2012), *Sentiment Analysis and Opinion and Opinion Mining*, Morgan & Claypool Publishers, 8
- [5] Cortes, C., & Vapnik, V. (1995). Support vector machine. *Machine learning*, 20(3), 273-297
- [6] Ken Schwaber y Jeff Sutherland (2017), *La Guía de Scrum*, 6,15
- [7] Gladwell Academy. <https://www.gladwell.nl/wat-is-scrum/>

8. Alumnos y Directores

Marcos Mauricio Carpintero Mendoza. - Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta: 2017630231, Tel. 5545495955, email: marcos_c79@hotmail.com

Firma: _____

Karla Alexia Reyes Reyes. - Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta: 2014101416, Tel. 5539356574, email: karr198854@gmail.com

Firma: _____

Olga Kolesnikova. - Doctora en Ciencias de Computación por el Centro de Investigación en Computación del IPN, Profesora de la ESCOM desde 2013 en el Departamento de Formación Básica. Áreas de investigación: procesamiento de lenguaje natural, inteligencia artificial. email: kolesolga@gmail.com.

Firma: _____

CARÁCTER: Confidencial
FUNDAMENTO LEGAL: Artículo 11 Fracc. V y Artículos 108, 113 y 117 de la Ley Federal de Transparencia y Acceso a la Información Pública.
PARTES CONFIDENCIALES: Número de boleta y teléfono.