**Algoritmo para el análisis de sentimientos de comentarios de clientes sobre productos de computación de Linio Colombia.**

**Johan Eduardo Cala Torra**

**Rober Julián Quintero Oviedo**

**Kevin Orlando Carreño Torres**

**Estructura de Datos**

**Matemáticas Discretas II**

**Lenin Javier Serrano Gil**

**Julián Darío Miranda Calle**

**UNIVERSIDAD PONTIFICIA BOLIVARIANA**

**ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA**

**BUCARAMANGA**

**2020**

1. **SITUACIÓN PROBLEMA**

En la actualidad, debido a la pandemia que se ha venido presentando mundialmente, y a consecuencia del distanciamiento social, la población se ha beneficiado del uso de tecnologías, empleándolas para realizar actividades las cuales ya no se pueden realizar presencialmente, como la compra de artículos por internet o mercado virtual, el cual es un tipo de comercio que se usa para realizar las transacciones de productos y servicios por medios electrónicos, en donde el uso de las tecnologías influye en la experiencia de compra del consumidor, por lo que las tiendas se convierten en lugares donde experimentar y compartir con otros consumidores sus opiniones.

Además de la experiencia de compra, algunas plataformas optan por eliminar la opción de comentarios y reseñas de los usuarios, temerosos de que las malas criticas o valoraciones afecten directamente al producto, igualmente es correcto afirmar que los comentarios positivos mejoran las cifras de ventas, pero para que ocurra esto es indispensable aceptar todo tipo de comentarios, por tal razón surge analizar los comentarios según los sentimientos positivos o negativos respecto a los productos.

Por lo tanto, surge la pregunta de investigación, ¿De qué manera se puede analizar los sentimientos de los usuarios al momento de comprar un producto de computación por mercado virtual?

1. **ESTADO DEL ARTE**
2. **OBJETIVOS**

**3.1 Objetivo General**

Desarrollar un algoritmo fundamentado en el procesamiento del lenguaje natural (NLP) para el análisis de sentimientos positivos y negativos que permitan el registro de los artículos que más atraen y menos atraen basados en sentimientos a través de los de comentarios de clientes sobre productos de computación de Linio Colombia.

**3.2 Objetivos Específicos**

3.2.1. Identificar los requerimientos funcionales y no funcionales del proyecto.

3.2.2. Describir técnicas para la adquisición y procesamiento de datos mediante procedimientos automatizados.

3.2.3. Diseñar modelos entidad-relación y diagramas UML.

3.2.3. Determinar el modelo de datos a partir de las técnicas de adquisición, a utilizar para el aprendizaje de procesamiento de lenguaje natural

3.2.4. Desarrollar el algoritmo que permita el procesamiento del lenguaje natural implícito en los comentarios de usuarios, para analizar los sentimientos positivos o negativos.

1. **JUSTIFICACIÓN**

El comercio virtual actualmente se ha favorecido, debido a la alta demanda de usuarios que se han ido incorporando, esto se debe a la situación actual de pandemia que se está viviendo, puesto que se debe prevenir el contacto físico, por tal razón muchas empresas han sido favorecidas vendiendo por este medio, gracias a las Tecnologías de la Información y la Comunicación, y al internet.

Este mercado puede tener ventajas, debido al alcance pues varios consumidores globalmente pueden acceder a las tiendas virtuales, además del tiempo de estar presente, están disponibles las 24 horas del día, pero a consecuencia de que el alcance es amplio, los consumidores pueden interactuar entre sí, es decir hay una relación entre los consumidores y los productos, en cuestión de segundos los usuarios pueden comentar opiniones positivas o negativas, hacer preguntas o quejarse sobre estos productos, lo que si no es correctamente manejado, puede ser una desventaja, debido a que si estos comentarios son negativas, el producto tiende a bajar en ventas, pero si es lo contrario podría incrementar sus compras respectivamente.

Por tal razón es importante analizar los comentarios, ya sean positivos o negativos de los usuarios, para poder que las empresas y la plataforma virtual pueda identificar la gran demanda de peticiones de los usuarios, para la detección de anomalías y la identificación de sentimientos en las peticiones y lograr darles una solución efectiva.

El Procesamiento de Lenguaje Natural (NLP), es una rama de la ciencia de datos que ayuda a las computadoras a entender, interpretar y manipular el lenguaje humano, NLP toma elementos prestados de muchas disciplinas, incluyendo la ciencia de la computación y la lingüística computacional, para interpretar y midan el sentimiento y determinen qué partes son importantes.

En nuestro proyecto es de suma importancia identificar y analizar los sentimientos de los comentarios, en la que el NLP pueda permitir el análisis de sentimientos de comentarios positivos o negativos de clientes sobre equipos de computación en la plataforma de comercio virtual Linio Colombia.

1. **METODOLOGÍA**

**La metodología Scrum**

Es una metodología ágil y flexible, permite en cualquier momento realinear el software con los objetivos, ya que puede introducir cambios funcionales o de prioridad en el inicio de cada nueva iteración sin ningún problema.

Debido a que es un proyecto en el cual se trabajara durante el semestre, en donde se deberán entregar avances, los cuales serán las entregas parciales para llegar a la entrega final del algoritmo fundamentado en NLP.

Una característica de la metodología Scrum, es que se ejecutaran actividades en ciclos temporales cortos y de duración fija, las cuales se realizaran semanales, hasta un máximo de 15 o 20 días si se requiere.

El proyecto va a tener una duración aproximada de 3 meses o menos, en los cuales tendrán que hacer las siguientes entregas:

* Primer avance del proyecto
* Artículo del proyecto 50%
* Avance del proyecto 50%
* Proyecto final 100%
* Artículo del proyecto 100%

Para realizar cada una de las anteriores entregas se desarrollarán las siguientes fases de la metodología Scrum (Figura 1), para definir actividades y entregar cada lunes en reuniones con el Scrum máster para ser revisados y evaluados, y finalmente realizarle la entrega al cliente en el tiempo establecido en el cronograma.

1. **Proceso de requisitos:**

En este proceso se identificarán los requisitos mínimos del algoritmo, como los requerimientos funcionales y no funcionales de nuestro algoritmo.

1. **Gestión de backlog:**

En este proceso se establecerán las actividades a realizar, las cuales estarán descritas en el cronograma (Sección 6) , donde serán establecidas las tareas por realizar para desarrollar el algoritmo.

1. **Planificación del sprint:**

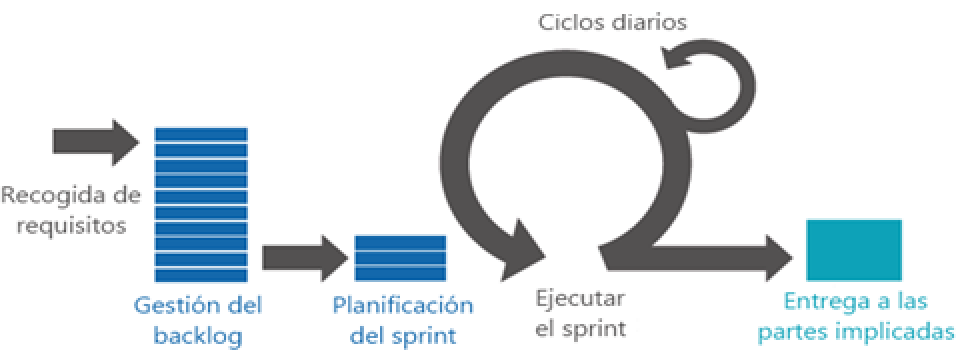
En este proceso el día de la reunión, se establecerá el tiempo y las actividades que se realizaran durante ese periodo, el cual podrá ser de 8 a 15 o más días si se requiere, luego se le asignara a cada integrante del equipo, se llevara un control por medio del tablero de Trello para Scrum.

1. **Ejecución del Sprint:**

Luego de terminar el periodo del Sprint, se realizará una reunión de sincronización y de acuerdo con lo establecido en el tablero de Trello, el Scrum Máster se encargará de inspeccionar que se estén realizando las actividades.

1. **Inspección e iteración**

El equipo presentara al cliente, los avances de proyecto en las fechas establecidas, en donde los docentes encargados, revisaran el entregable con sus respectivos comentarios, para así luego poder adaptarse para lograr tener el algoritmo planteado.



*Figura 1. Fases metodología Scrum tomada de https://www.ida.cl/blog/estrategia-digital/metodologia-scrum-en-proyectos-digitales/*

1. **CRONOGRAMA DE ACTIVIDADES**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ACTIVIDADES** | **DURACIÓN (SEMANAS)** | **FECHA  INICIO** | **FECHA  FIN** | **ENTREGABLES** | **ENTREGABLE FINAL** |
| Actividad 1.1: Documentación de la propuesta | 2 | Julio 17 de 2020 | Julio 31 de 2020 | Documento de la propuesta del proyecto | **Propuesta del proyecto** |
| Actividad 1.2: Realizar acta de requerimientos funcionales, no funcionales y restricciones de software | 1 | Agosto 14 de 2020 | Agosto 25 de 2020 | Acta de requerimientos funcionales, no funcionales y restricciones de software | **Primer Avance del proyecto** |
| Actividad 2.1: Investigar sobre técnicas para la adquisición de datos, dando claridad la procedencia de los datos que serán usados para el entrenamiento de modelos para el NPL | 1 | Agosto 14 de 2020 | Agosto 25 de 2020 | Documentación sobre la adquisición de datos |
| Actividad 2.2: Describir las actividades iniciales del preprocesamiento y estandarización de los datos | Agosto 17 de 2020 | Agosto 20 de 2020 |
|  |
| Actividad 3.1: Realizar el modelo entidad - relación de la base de datos, con su diccionario de datos | 1 | Agosto 20 de 2020 | Agosto 25 de 2020 | Documentación modelo entidad - relación con su diccionario de datos |  |
| Actividad 3.2.: Realizar los diagramas de clases, componentes y actividades para el diseño del algoritmo | Agosto 22 de 2020 | Agosto 25 de 2020 | Documentación diagramas de clases, componentes y actividades |  |
| Actividad 4.1: Realizar la adquisición de datos, obteniéndolos a partir de las técnicas automatizadas investigadas. | 1 | Agosto 26 de 2020 | Septiembre 6 de 2020 | Documentación de la base de datos local con Data Wragnling | **Avance del proyecto 50%** |  |
| Actividad 4.2: Modelar los datos adquiridos, en una base de datos local, preparando los datos Data Wragnling | 2 | Septiembre 7 de 2020 | Septiembre 18 de 2020 | Base de datos local |  |
| Actividad 5.1: Investigar sobre técnicas de Machine Learning, para el procesamiento del lenguaje natural. | 2 | Septiembre 7 de 2020 | Septiembre 18 de 2020 | Documentación sobre técnicas implementada para Machine Learning |  |
| Actividad 5.2: Desarrollar el algoritmo usando técnicas de Machine Learning, para el procesamiento del lenguaje natural NLP | 3 | Septiembre 14 de 2020 | Octubre 2 de 2020 | Algoritmo fundamentado en el procesamiento del lenguaje natural | **Proyecto final 100%** |  |
| Actividad 5.3: Implementar la base de datos local en el desarrollo del algoritmo | 1 | Septiembre 21 de 2020 | Septiembre 28 de 2020 | Algoritmo implementando la base datos de manera local |  |
| Actividad 5.4: Validar el correcto funcionamiento del algoritmo, teniendo en cuenta los requerimientos planteados. | 1 | Octubre 2 de 2020 | Octubre 6 de 2020 | Algoritmo final para el análisis de sentimientos de comentarios de clientes |  |
| Actividad 5.5: Realizar la documentación del algoritmo, el artículo de investigación. | 3 | Octubre 5 de 2020 | Noviembre 2 de 2020 | Artículo de investigación, documentación final del algoritmo | **Artículo de investigación** |  |

**REFERENCIAS BIBLIOGRÁFICAS**

[1]G. Carlos and G. Sonia, "Sistema Deep Learning para el análisis de sentimientos en opiniones de productos para la ordenación de resultados de un buscador semántico", Hdl.handle.net, 2020. [En línea]. Available: http://hdl.handle.net/2183/25152. [Último acceso: 31 Julio 2020].

[2]A. Tripathy and S. Rath, "Classification of Sentiment of Reviews using Supervised Machine Learning Techniques", International Journal of Rough Sets and Data Analysis, vol. 4, no. 1, pp. 56-74, 2017. Available: 10.4018/ijrsda.2017010104.

[3]G. Carlos and G. Sonia, "Sistema Deep Learning para el análisis de sentimientos en opiniones de productos para la ordenación de resultados de un buscador semántico", Hdl.handle.net, 2020. [En línea]. Available: http://hdl.handle.net/2183/25152. [Último acceso 31 Julio 2020].

[4]P. Baid, A. Gupta and N. Chaplot, "Sentiment Analysis of Movie Reviews using Machine Learning Techniques", International Journal of Computer Applications, vol. 179, no. 7, pp. 45-49, 2017. Available: 10.5120/ijca2017916005.

[5]ProyectoAgiles.org., «Delaossa,» 23 Agosto 2016. [En linea]. Available: https://delaossa.co/blog/36-proyecto-digital/143-la-metodologia-scrum-al-servicio-de-los-proyectos-digitales. [Último acceso: 25 Julio 2020].

[6] SAS, «SAS,» [En línea]. Available: https://www.sas.com/es\_co/insights/analytics/what-is-natural-language-processing-nlp.html. [Último acceso: 21 agosto 2020].

[7] E. Higuerey, «rocktcontent,» [En línea]. Available: https://rockcontent.com/es/blog/comercio-electronico/. [Último acceso: 20 agosto 2020].

**FICHA DE RESUMEN ANALÍTICO**

*0. Pregunta de investigación*

¿Cómo podemos analizar los sentimientos de los clientes sobre los productos de computación de Linio Colombia a través del Procesamiento del Lenguaje Natural?

*1. Ficha bibliográfica*(copiar del formato correspondiente de la base de datos bibliográficos)

[1]G. Carlos and G. Sonia, "Sistema Deep Learning para el análisis de sentimientos en opiniones de productos para la ordenación de resultados de un buscador semántico", Hdl.handle.net, 2020. [Online]. Available: http://hdl.handle.net/2183/25152. [Accessed: 31- Jul- 2020].

*2. Problemática de investigación / Objeto de estudio/Estudios previos realizados:*

-Análisis de Sentimientos, enmarcado dentro del área de estudio del Procesamiento de Lenguaje Natural

*3. Datos y análisis/ descripciones relevantes*

-la aplicación de los algoritmos de aprendizaje clásicos, para establecer una base, para posteriormente utilizar métodos de aprendizaje automáticos como Deep Learning para que sean capaces de aprender a discernir la estructura de las oraciones y así mejorar ampliamente los resultados.

*4. Nuevos aportes realizados en el artículo*

- Escriba aquí cuáles son los aportes (o trascendencia) que deja el artículo leído. Esto es, revisar detalladamente sus ventajas y limitaciones.

*5. Resultados / conclusiones:*

Se llego a la conclusión que las topologías complejas tienen tendencia a adaptarse mas rápido a los conjuntos a los conjuntos de entrenamiento provocado un sobre entretenimiento, esto se debe a la simplicidad del conjunto de datos utilizados.

*7. Aportes a la investigación:*

- Debemos tomar este trabajo como un gran ejemplo a seguir, ya que su el objetivo de este se asemeja a nuestra pregunta, se vio el aporte del Deep Learning un método que aunque mas complejo, nos puede ser de gran ayuda para lograr un mejor resulltado

Fecha de elaboración: Día: 22\_ Mes: 07\_ Año:\_\_\_2020\_\_\_

Responsable: Johan Cala T.