

Revisión del análisis de sentimientos con opiniones en español de hoteles de Medellín

Santiago Quintero Hincapié, Daniel Santa Rendón, Daniel Torres González

*Ingeniería de sistemas, Universidad de Antioquia
Medellín, Colombia*

santiago.quintero@udea.edu.co

daniel.santar@udea.edu.co

daniel.torresg@udea.edu.co

Resumen

El análisis de sentimientos es una herramienta que está ganando cada vez más popularidad por sus grandes ventajas y beneficios de uso que consiste en clasificar un texto como positivo, negativo o neutro. Este se ha convertido en una herramienta muy útil para analizar la percepción acerca de un tema, producto o servicio, plasmada en un texto escrito. Para muchas entidades, tanto del sector privado como del sector público, contar con una herramienta que pueda analizar y clasificar de forma automática como positivo, negativo o neutro, la gran cantidad de información que se genera en torno a sus marcas, servicios o productos, se ha transformado en un componente indispensable al momento de tomar decisiones corporativas. En este artículo se puede encontrar un estimado de las cantidades de trabajos de este tema que han utilizado bases de datos en español e inglés, con el fin de demostrar la falta de investigación que hay sobre el análisis de sentimientos a nivel de documento en español, particularmente en el campo de las opiniones de hoteles en Medellín. Además, se listan algunas empresas que prestan el servicio de análisis de sentimientos y las organizaciones que utilizan estos servicios para dar una idea de cual es el interés del sector empresarial actualmente sobre el análisis de sentimientos.

El contexto principal de este proyecto es académico, pero también es organizacional, específicamente enfocado en hoteles de Medellín. Es académico debido a que es una investigación fundamental en la que se quiere examinar los métodos existentes para el análisis de sentimientos con opiniones en español de hoteles, pero también es aplicada y enfocada a las organizaciones puesto que se busca aplicar el método más efectivo para hacer análisis de sentimientos en hoteles de la ciudad de Medellín. Para abordar la solución se propone evaluar los métodos para el análisis de sentimientos y, así, conocer el mejor de estos para las opiniones de hoteles en español.

Palabras claves

Análisis de sentimientos, opiniones, hoteles, hoteles de Medellín, español, reviews, bases de datos.

1. Introducción

El análisis de sentimientos, también denominado minería de opiniones, es el campo de estudio que analiza las opiniones, sentimientos, evaluaciones, valoraciones, actitudes y emociones de las personas hacia entidades como productos, servicios, organizaciones, individuos, temas, eventos y sus atributos (Bing Liu. 2012).

Este artículo tiene la intención de argumentar porqué los métodos de análisis de sentimientos con opiniones de hoteles en español es un campo de investigación no explorado e implementado a profundidad en Colombia, específicamente en Medellín. Como el contexto de esta investigación está más inclinado al lado de lo fundamental, la perspectiva del trabajo es científica, es decir, se busca dar respuesta a un problema de conocimiento.

La investigación se hace por medio de una revisión sistemática de literatura de los últimos 5 años, en primer lugar, de artículos que usan bases de datos en inglés contra artículos que usan bases de datos en español para aplicar métodos de análisis de sentimientos; luego se reduce la búsqueda a artículos que usan bases de datos en inglés comparado con artículos que usan bases de datos en español de opiniones de hoteles; en tercer lugar, se tratará de encontrar artículos que realicen análisis de sentimientos con opiniones en español de hoteles de Medellín, además, se investigará empresas prestadoras del servicio de análisis de sentimientos y se hará un análisis de los clientes de éstas para dar un panorama un poco más amplio de los intereses organizacionales.

Las siguientes secciones incluyen un repaso del problema, el contexto, las causas y consecuencias del problema, y luego, una revisión de

En la sección 2 de este artículo se plantea el problema, el contexto, un marco teórico, causas y consecuencias del problema, para luego empezar con una revisión sistemática de literatura para obtener suficiente información con la cual se pueda sustentar el problema tratado. En la sección 4 se exponen algunas soluciones que han existido para problemas similares, acompañadas de sus pros y sus contras. Por último, se plantean las conclusiones encontradas del trabajo realizado en la sección 4.

2. Situación problemática

El mundo está avanzando día a día a pasos agigantados en comparación a épocas anteriores por medio de la investigación. Partiendo desde este punto, se debe tener en cuenta que para cada uno de los avances investigativos debe existir una documentación que mantenga en registro el desarrollo de tal avance, este registro, tiene como ideal principal poder transmitir el mensaje lo más rápido, claro y fácil posible. Una de las alternativas usadas para que estos criterios se cumplan o, por lo menos, se acerquen más al objetivo de comunicarlo es escribir en inglés, y como este es el idioma universal se han hecho muchas más aplicaciones y experimentos utilizando esta lengua, con el fin de llegar a una mayor cantidad de personas. Aplicado en el contexto de este artículo, existen trabajos realizados con bases de datos de opiniones en español de análisis de sentimientos pero comparados con las investigaciones con bases de datos de opiniones en inglés, estos son una minoría. Otra causa del problema es que las empresas colombianas apenas están incursionando en el mundo del análisis de sentimientos por lo que no hay una inversión considerable del sector industrial en este campo; un posible elemento de esta causa es la falta de enseñanza sobre el análisis de sentimientos en las universidades (esto último queda como supuesto ya que no se investigará en este artículo).

Como consecuencia de esto, no existe la suficiente documentación para saber cual método es el más adecuado para realizar análisis de sentimientos dentro del contexto de opiniones en español de hoteles, particularmente, hoteles de la ciudad de Medellín.

El análisis de sentimientos, también denominado minería de opiniones, es el campo de estudio que analiza las opiniones, sentimientos, evaluaciones, valoraciones, actitudes y emociones de las personas hacia entidades como productos, servicios, organizaciones, individuos, temas, eventos y sus atributos (Bing Liu. 2012).

El análisis de sentimientos ya ha sido abordado en diferentes niveles tales como: a nivel de documentos, a nivel de oraciones y a nivel de entidades y aspectos, este último llamado también nivel de características (Hu and Liu. 2004).

Los documentos de opinión no solo existen en la web (a menudo llamado datos externos); muchas organizaciones

tienen datos internos, como por ejemplo, *feedback* de clientes recolectados de e-mails y call centers y encuestas conducidas por organizaciones (Bing Liu. 2015), es de vital importancia analizar estos tipos de datos para así poder resumir la opinión de los usuarios.

En este proyecto nos centraremos en el análisis de sentimientos a nivel de opiniones de hoteles en español. Sabiendo que el análisis de sentimientos a este nivel es quizás el más extenso tema de estudio en el campo del análisis de sentimientos, especialmente en sus primeros días (Surveys by Pang and Lee, 2008; Liu, 2012). Y sabiendo también que un documento está compuesto por párrafos y oraciones, inicialmente la tarea es determinar si estos expresan una opinión o sensación positiva, negativa o neutra, aunque también existen otros dos tipos de opiniones, opiniones regulares y opiniones de comparación, lo cual pone las cosas aún más interesantes y retadoras (Jindal and Liu. 2006b). Este nivel de análisis está altamente relacionado con la clasificación de subjetividad (Wiebe et al. 1999). Se discutirá los principales temas de investigaciones del análisis de sentimientos y su estado del arte en lo concerniente a algoritmos.

Otras investigaciones también han analizado las frases (Wilson et al. 2004), pero el nivel de la cláusula aún no es suficiente, por ejemplo, "*Apple* lo está haciendo muy bien en esta pésima economía." en este nivel de análisis implícitamente se asume que cada párrafo expresa una opinión en una única entidad como por ejemplo un único producto o servicio, así que no es aplicable a documentos que evalúan o comparen múltiples entidades, para los cuales es necesario un análisis más detallado (Bing Liu. 2015). Por otra parte se tiene el concepto de subjetividad que ha sido ampliamente usado en análisis de sentimientos. Este concepto ha causado algunas confusiones entre investigadores. En muchos artículos, ser subjetivo y ser poseedor de sentimientos son considerados como equivalentes, pero no son lo mismo. Una oración objetiva establece alguna información objetiva, mientras una frase subjetiva expresa algún sentimiento personal, punto de vista, juicio, o creencias. Un ejemplo de una frase objetiva es "*The iphone is an Apple product*", un ejemplo de una frase subjetiva es "*I like the iphone*" la tarea es determinar ya sea si una oración es subjetiva u objetiva, esto es llamado clasificación de subjetividad (Wiebe Janyce and Ellen Riloff. 1999).

La manera correcta de abordar este análisis de sentimientos es identificando palabras y modismos que demuestran una opinión o sensación dentro de una oración, esta lista de palabras y modismos es llamada léxico de sentimientos, sin embargo, aunque ésta es necesaria para realizar un correcto análisis de sentimientos, no es suficiente (Bing Liu. 2012).

En general, dado un grupo de opiniones, el análisis de sentimientos consistirá en lo siguiente: extracción y categorización de entidades, aspectos, opinión del titular, fecha de la opinión dada y sentimiento de la opinión (Liu. 2006, 2011). Esta clasificación se puede llevar a cabo con algoritmos de *machine learning*, debido a que la clasificación de sentimientos se reduce a un problema de clasificación de texto, algunos métodos existentes de

aprendizaje supervisado pueden ser directamente aplicados, tal como clasificación de *naïve Bayes* o *support vector machines* (SVM) (Joachims, 1999; Shawe-Taylor and Cristianini, 2000).

2.1. Búsqueda

Para realizar una búsqueda rigurosa de información se plantean los siguientes elementos:

2.1.1. Bases de datos

Las bases de datos usadas para la búsqueda de información serán las ofrecidas por la universidad:

Principales: Bases de datos a las que se les aplicará el método de revisión sistemática de literatura.

- IEEE xlore
- ScienceDirect

Se eligieron estas dos como principales, inicialmente por el trabajo de revisar cada artículo de cada base de datos y también porque son referentes para la ingeniería.

Secundarias: Bases de datos en las cuales solo se realizará la búsqueda de las cadenas propuestas más adelante.

- Scielo
- Scopus
- SpringerLink

2.1.2. Criterios de inclusión y exclusión

Se plantean los siguientes criterios de inclusión y exclusión para los artículos encontrados. Cada tabla de criterios hace referencia a cada una de las búsquedas planteadas en este artículo.

1. Artículos que usan métodos de análisis de sentimientos con bases de datos en inglés ó español:

<i>Criterios de inclusión</i>	<i>Criterios de exclusión</i>
Artículos publicados en 2014 o después, y	Artículos publicados antes del 2014, o
Que implementen algún método de análisis de sentimientos usando bases de datos en inglés ó en español , y	Que no hagan uso de bases de datos en inglés ó español para el entrenamiento de métodos de análisis de sentimientos, o
Que estén publicados en bases de datos elegidas.	Que estén publicados en fuentes diferentes a las seleccionadas.

2. Artículos que usan métodos de análisis de sentimientos con bases de datos en inglés ó español de opiniones de hoteles:

<i>Criterios de inclusión</i>	<i>Criterios de exclusión</i>
Artículos publicados en 2014 o después, y	Artículos publicados antes del 2014, o
Que implementen algún método de análisis de sentimientos usando bases de datos en inglés ó en español puntualmente de opiniones de hoteles , y	Que no hagan uso de bases de datos en inglés ó español para el entrenamiento de métodos de análisis de sentimientos usando bases de datos de opiniones de hoteles , o
Que estén publicados en bases de datos elegidas.	Que estén publicados en fuentes diferentes a las seleccionadas.

3. Artículos que usan métodos de análisis de sentimientos con bases de datos de opiniones en español de hoteles de la ciudad de Medellín:

<i>Criterios de inclusión</i>	<i>Criterios de exclusión</i>
Artículos publicados en 2014 o después, y	Artículos publicados antes del 2014, o
Que implementen algún método de análisis de sentimientos usando bases de datos de opiniones en español de hoteles de la ciudad de Medellín , y	Que no apliquen métodos de análisis de sentimientos o que no hagan uso de bases de datos de opiniones en español de hoteles en la ciudad de Medellín, o
Que estén publicados en bases de datos elegidas.	Que estén publicados en fuentes diferentes a las seleccionadas.

2.1.3. Método para la revisión bibliográfica

Se propone una adaptación del método de revisión sistemática de literatura propuesto por (Jaramillo, 2015). Este consiste en que en base a unos criterios de inclusión y de exclusión se filtran los trabajos; primer paso: se lee el título, el abstract y si es necesario la introducción y conclusiones, por último, se lee la sección experimental para conocer qué tipo de bases de datos usaron en sus pruebas. Segundo paso o proceso de bola de nieve: por medio de los criterios antes mencionados, se seleccionan las referencias que cumplan con todos los requisitos de inclusión y se comienza a repetir los pasos de la primera etapa con las referencias de estos documentos y se evalúan con los mismos criterios; el límite seleccionado en el proceso de bola de nieve es de dos referencias desde el artículo original.

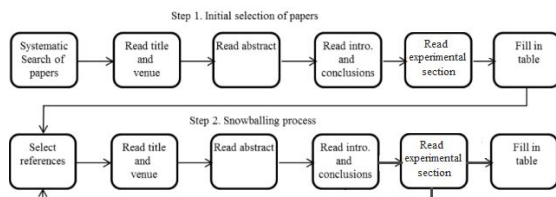


Fig 1. Método para la revisión de artículos (Jaramillo, 2015)

2.1.4. Cadenas de búsqueda

En algunas bases de datos, el símbolo “+” debe ser cambiado por AND, pero, por convención, para referirnos a esto, únicamente utilizaremos el símbolo “+”.

Para la primera búsqueda:

- “sentiment analysis” + english
- “sentiment analysis” + spanish

Para la segunda búsqueda:

- “sentiment analysis” + hotel
- “sentiment analysis” + hotel + spanish
- “sentiment analysis” + hotel + english

Para la tercera búsqueda:

- “sentiment analysis” + Medellin hotel
- “sentiment analysis” + Medellin hotel + english
- “sentiment analysis” + Medellin hotel + spanish

2.2. Resultados

Es importante aclarar que por falta de tiempo y cantidad de información no fue posible leer todos los resultados encontrados en las base de datos, sin embargo es un estimado de la cantidad de artículos.

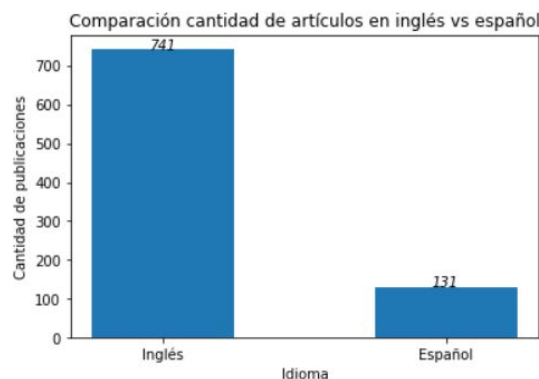
2.2.1. Primera búsqueda

Resultado parcial de artículos encontrados que emplean métodos de sentimientos a bases de datos en inglés vs bases de datos en español

Para las bases de datos principales, no se pudo realizar una búsqueda completa y detallada por la cantidad de artículos y lo complicado de clasificar cuáles de estos usan bases de datos en inglés, por esto se da un aproximado con la cadena de búsqueda “*sentiment analysis*” + *english* y que se incluyan dentro de las categorías de *ciencias de la computación*, *ingeniería*, *conocimiento de bases de datos*, *sistemas expertos con aplicación*, *procesamiento y manejo de información*, y *análisis e investigación de turismo*. Con estas categorías, 186 artículos fueron encontrados en IEEE y 555 en ScienceDirect, si es solo “*sentiment analysis*” se encuentran 3453 en IEEE y 3116 en ScienceDirect, de los cuales la mayoría usa bases de datos en inglés pero al no ser explícito en el artículo, el proceso de filtrado se vuelve muy complejo.

Usando	sentiment analysis + english	sentiment analysis + spanish
IEEE xplora	186	21
Scielo	2	2
ScienceDirect	555	110
Scopus	159	50
SpringerLink	19	4
Total	921	187

En las siguientes gráficas únicamente se tomaron en cuenta los artículos encontrados en las bases de datos principales, es decir, IEEE xplora y ScienceDirect.

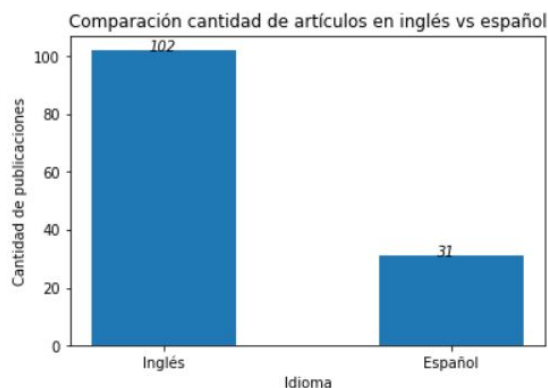


En todas las bases de datos se observa un comportamiento similar, el número de artículos que usan bases de datos en inglés es mucho mayor al número de artículos que usan bases de datos en español. Confirmando, no la ausencia de documentación sobre análisis de sentimientos a nivel del español, pero si la insuficiencia que hay con respecto al inglés.

2.2.2 Segunda búsqueda

Usando	sentiment analysis + hotel + english	sentiment analysis + hotel + spanish
IEEE xplora	7	0
Scielo	0	0
ScienceDirect	95	31
Scopus	6	6

SpringerLink	3	1
Total	111	38



Esta búsqueda al ser más específica, los resultados fueron mucho menores que la búsqueda anterior, aún así se sigue el mismo comportamiento, excepto en Scielo donde no se encontró información del tema y Scopus que arrojaron la misma cantidad de resultados para análisis de sentimiento con opiniones en español de hoteles y para análisis de sentimiento con opiniones en inglés de hoteles.

Links para ver algunos de los artículos encontrados en [español](#) e [inglés](#).

2.2.3. Tercera búsqueda

Dentro de las temáticas ingeniería y ciencia de la computación, no se encontró ningún artículo en las bases de datos usando las cadenas:

- sentiment analysis + Medellín hotel + english
- sentiment analysis + Medellín hotel + spanish

Esto demuestra la nula existencia de investigaciones en los hoteles del valle de Aburrá sobre el análisis de sentimientos.

2.2.4. Empresas que prestan los servicios de análisis de sentimientos

SentiMetrix (<http://www.sentimetrix.com/>):

Las organizaciones a las que presta servicios son:

- Tata consultancy services
- Forbes.com
- Rensselaer
- Stanford University
- The world bank
- Darpa
- Center for international rehabilitation
- U.S Army
- Farm Credit Canada
- SAATCHI & SAATCHI

Hi-Tech BPO(<https://www.hitechbpo.com/>):

Las organizaciones a las que presta servicios son:

- SIEMENS
- Volkswagen
- Domino's pizza
- Dow International B.V.
- Westfield
- Monash University
- KFORCE
- STEVENS
- TIBCO
- Griffith University
- Pfizer

Sentisum (<https://www.sentisum.com/>):

Las organizaciones a las que presta servicios son:

- DHL
- REEVOO
- Infinity
- British Airways
- AXA

socialbakers (<https://www.socialbakers.com/>):

Las organizaciones a las que presta servicios son:

- DANONE
- Jägermeister
- OgilvyOne
- Y&R
- FrieslandCampina
- McDonald's
- TRUSSARDI
- EUROSPORT
- La Redoute
- ASUS
- TEB
- LEXUS
- TOYOTA
- NESTLÉ

Brand24 (<https://brand24.com/>):

Las organizaciones a las que presta servicios son:

- OLX
- watershed
- zortrax
- IKEA
- ORBITVU
- VICHY

inbenta (<https://www.inbenta.com/es/>):

- Claro
- ICETEX
- skyscanner
- movistar
- drchrono
- REPSOL
- RAPP
- PROSEGUR
- ICFES
- IBERIA
- BBVA

La gran mayoría de organizaciones que adquieren un servicio de análisis de sentimiento son Europeas,

Estadounidenses y Asiáticas. Muy pocas empresas latinas están involucradas en el análisis de sentimientos, lo que significa una oportunidad de explorar el campo industrial y gubernamental de latinoamérica, siendo los hoteles de Medellín el contexto inicial para la futura investigación.

2.2.5. Reviews

“Every customer engagement is an opportunity to enhance our relationship and create a great experience for our community. That’s why Socialbakers is an integral part of our community management strategy. With Socialbakers’ easy-to-use interface, we have the ability to close the gap in cross-team communication so we can give our customers the attention they deserve.” Osmat Awar, Director, Corporate Communication and Social Responsibility McDonald’s RICC

“Socialbakers enables us to understand our social channels’ competitive status in the shortest time – under one dashboard. We’re able to monitor what’s generating the most interactions quickly, and can spend more time on creating great content instead of on gathering the data.” Jeff Lee, Head of Community and Online Media Buy

“At an overarching brand level, SentiSum allows us to set targets and monitor success measures on the key pillars that underpin our ambition. We can identify the drivers behind our NPS scores. In less than 5 minutes, we are now able to understand the drivers of our advocacy from over 100k reviews” Head of Customer Service & Quality, British Airways Holidays

Estas reviews de empresas demuestran que las herramientas de análisis de sentimientos son una gran ayuda al momento de conocer las opiniones de los clientes, en qué mejorar, en qué productos se interesan más, etc. Corroborando así, la oportunidad de negocio en un sector poco explorado como lo es la industria hotelera de Medellín.

3. Discusión

El análisis de sentimientos en opiniones de hoteles ha sido abordado en otros idiomas con resultados interesantes, como por ejemplo en el caso de Lisa Medrouk y Anna Pappa quienes trabajaron particularmente en reseñas de hoteles y restaurantes que están escritas en diferentes idiomas tales como francés, inglés y griego. En este caso, ellas para el enfoque de aprendizaje automático, utilizaron dos redes neuronales profundas diferentes, Redes neuronales convolucionales (ConvNets) y Redes neuronales recurrentes (RNN).

El modelo de aprendizaje que ellas aplicaron explota la información a nivel de n-gramas y logra una alta precisión para la polaridad de los sentimientos y la clasificación de temas de acuerdo con las pruebas y resultados experimentales. teniendo en cuenta sus resultados, ellas tienden a promover el inicio de un enfoque simple pero poderoso para alimentar redes

profundas en un contexto multilingüe. (Medrouk y Pappa, 2018).

Otra solución es la planteada por Xiaobo Zhang y Qingsong Yu, en la cual plantean el uso de un vector de características reducido por medio de la herramienta Word2Vec y el algoritmo de agrupación ISODATA. Emplear esta herramienta y este algoritmo trae ventajas destacables como la posibilidad de usar palabras únicas y de evitar la necesidad de conocer mucho de la base de datos. Por otro lado, implementar Word2Vec podría provocar que el sentido de las palabras no se capture de manera independiente, por ejemplo, dos palabras escritas iguales con más de un sentido, serían clasificadas como iguales, así mismo, al combinarse con el algoritmo ISODATA, podrá llevar mucho tiempo si la base de datos es muy desestructurada, ya que por sí solo, tiende a tardar. (Zhang, X., & Yu, Q. 2017)

Por otra parte, Qianjun Shuai *et al.* llevaron a cabo un análisis de opinión sobre las revisiones de hoteles sacadas de Doc2vec y algoritmos de machine learning como Support Vector Machine (SVM), Logistic Regression y Naïve Bayes. Evaluaron el rendimiento de estos algoritmos usando índices de precisión, tasa de recuperación y F-measure. En este caso el que mejor dio resultado fue el SVM. Proponen para futuras investigaciones, mejorar los resultados usando más datos o mezclar el vector de características y el diccionario de sentimientos. Además, también consideran el uso de algoritmos de *deep learning* como las redes neuronales convolucionales. (Shuai, Q., Huang, Y., Jin, L., & Pang, L. 2018)

Aunque estos métodos dan resultados positivos, hay que tener en cuenta, aparte de las contras que tiene cada método, que no han sido entrenados y validados con opiniones de hoteles en español, y mucho menos, de hoteles de Medellín. Es importante aclarar que cada idioma maneja unos modismos particulares los cuales hacen variar drásticamente el significado de una opinión y por lo tanto, el resultado del algoritmo.

4. Conclusiones

Podemos concluir que el análisis de sentimientos si es deseado por entidades que venden productos y servicios, si es efectivo y lo más importante es que si han tenido buenos resultados. Sumado a esto, gracias a la rigurosa búsqueda se logró identificar que, comparado con el inglés, la cantidad de artículos que utilizaron bases de datos en español es muy poca, por tal motivo existen muchos contextos en español en los que aún falta investigar e implementar inteligencia artificial para el análisis de sentimientos, en nuestro caso, el contexto de hoteles de Medellín.

Existen otras soluciones, pero que hacen uso de bases de datos en idiomas diferentes al español, por lo que no se puede decir, a simple vista, si el método es viable o no, debido a las variaciones de modismos y expresiones entre idiomas, países y regiones.

Referencias Bibliográficas

Bing Liu. Sentiment Analysis. Mining Opinions, sentiments and Emotions. 2015

Bing Liu. *Sentiment Analysis and Opinion Mining*. 2012

Chen, Yubo and Jinhong Xie. *Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix*. 2008

Hu, Minqing and Bing Liu. *Mining and summarizing customer reviews*. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)*. 2004.

Jaramillo, A. *Requirements Elicitation Approaches: A Systematic Review*. 2015

Jindal, Nitin and Bing Liu. *Mining comparative sentences and relations*. In *Proceedings of National Conf. on Artificial Intelligence (AAAI-2006)*. 2006b.

Liu, Bing. *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. 2006 and 2011: Springer.

Lisa Medrouk y Anna Pappa. *Do Deep Networks Really Need Complex Modules for Multilingual Sentiment Polarity Detection and Domain Classification?* 2018.

Nello Cristianini, John Shawe-Taylor. *An introduction to support Vector Machines: and other kernel-based learning methods*. 1999

Shuai, Q., Huang, Y., Jin, L., & Pang, L. *Sentiment Analysis on Chinese Hotel Reviews with Doc2Vec and Classifiers*. 2018

Surveys by Pang and Lee, Liu. *Sentiment Analysis and Opinion Mining*. 2012

Wiebe, Janyce, Rebecca F. Bruce, and Thomas P. O'Hara. *Development and use of a gold-standard data set for subjectivity classifications*. In *Proceedings of the Association for Computational Linguistics (ACL-1999)*. 1999.

Wiebe Janyce, Ellen Riloff. *Learning Extraction Patterns for Subjective Expressions*. 1999

Wilson, Theresa, Janyce Wiebe, and Rebecca Hwa. *Just how mad are you? Finding strong and weak opinion clauses*. In *Proceedings of National Conference on Artificial Intelligence (AAAI-2004)*. 2004.

Zhang, X., & Yu, Q. *Hotel reviews sentiment analysis based on word vector clustering*. 2017.