

# Chapter 35

## An Opinion Mining Approach for Drug Reviews in Spanish

**Karina Castro-Pérez**

 <https://orcid.org/0000-0002-1709-0087>

*Tecnológico Nacional de México, Mexico & IT Orizaba, Mexico*

**José Luis Sánchez-Cervantes**

 <https://orcid.org/0000-0001-5194-1263>

*CONACYT, Mexico & Instituto Tecnológico de Orizaba, Mexico*

**María del Pilar Salas-Zárate**

*Tecnológico Nacional de México, Mexico & ITS Teziutlán, Mexico*

**Maritza Bustos-López**

*Tecnológico Nacional de México, Mexico & Instituto Tecnológico de Orizaba, Mexico*

**Lisbeth Rodríguez-Mazahua**

*Tecnológico Nacional de México, Mexico & Instituto Tecnológico de Orizaba, Mexico*

### ABSTRACT

*In recent years, the application of opinion mining has increased as a boom and growth of social media and blogs on the web, and these sources generate a large volume of unstructured data; therefore, a manual review is not feasible. For this reason, it has become necessary to apply web scraping and opinion mining techniques, two primary processes that help to obtain and summarize the data. Opinion mining, among its various areas of application, stands out for its essential contribution in the context of healthcare, especially for pharmacovigilance, because it allows finding adverse drug events omitted by the pharmaceutical companies. This chapter proposes a hybrid approach that uses semantics and machine learning for an opinion mining-analysis system by applying natural-language-processing techniques for the detection of drug polarity for chronic-degenerative diseases, available in blogs and specialized websites in the Spanish language.*

DOI: 10.4018/978-1-6684-6303-1.ch035

## INTRODUCTION

Opinion mining is an area of great importance for the coarse application that has, focuses on analyzing opinions, sentiments, evaluations, assessments, attitudes, and emotions of people towards entities such as products, services, organizations, individuals, problems, and events (Liu, 2012; Jiménez et al., 2018). This technique, emerged thanks to the accelerated growth of resources available on the web, a representative work of the application, in its beginnings, of opinion mining is the study of Das and Chen (2004) in which they found that, in the case of Amazon Inc., there were cumulatively 70,000 messages by the end of 1998 on Yahoo's message board, and this had grown to about 750,000 messages by early 2004. The authors found that many of the messages from Amazon's board offered favorable, pessimistic, confusing, and even spamming opinions, so in their study, they demonstrated the possibility of capturing sentiment by applying statistical language and natural language processing techniques. Thus, sentiment analysis, also known as opinion mining, began to take on relevance. Also, the availability of potential resources for analysis continued to grow exponentially, now through online review sites, shopping sites, and blogs, which increased the challenge of understanding people's opinions, those opinions important to the decision-making process. Pang and Lee (2008) analyzed surveys of American adults, where they found that consumers reported being willing to pay 20% to 99% more for a 5-star item than a 4-star item in an online store, clearly identifying the importance of knowing other people's opinions and feelings about a product.

It is noteworthy that opinion mining is not only applied for analysis in the consumption of goods and services; it has also been applied in the field of politics; generally, it recognized by the full range of applications it has at present. On the other hand, to implement opinion mining, it is necessary to use lexical resources that help to carry out sentiment classification. A widely known resource is SENTIWORDNET 3.0 (Baccianella et al., 2014) created for research purposes, which provides automatic annotation of all WORDNET synsets according to their degrees of positivity, negativity, and neutrality. The process presented by the authors consists of two steps,

- 1) weak-supervision, semi-supervised learning step;
- 2) a random-walk step, used to support sentiment classification in opinion mining.

Thus, the industry surrounding sentiment has grown due to the proliferation of analytics for commercial applications, as well as the exponential increase, in recent years, of social networks and video blogs accessed by millions of users, which generate large amounts of unstructured data. Given this fact, manual revision of data is not very feasible, as a consequence of the number of data continuously generated; therefore, it highlights the use and implementation of opinion mining in current systems (Liu, 2012).

Notwithstanding, opinion mining in the area of health care has increased because of the benefits provided for decision-making, one of the analyses that can be performed using this technique is pharmacovigilance, defined as the science and activities related to the detection, assessment, understanding, and prevention of adverse effects or any other drug-related problem (*World Health Organization*, 2015).

In this context, according to the World Health Organization in 2016, diabetes mellitus, hypertension, cardiovascular diseases, cancer, among other diseases, classified as chronic-degenerative, are positioned among the top ten causes of death in Mexico and the world (World Health Organization, 2020). This means that health systems and experts need to analyze important aspects such as eating habits, exercise, and treatments. Therefore, the application of opinion mining is useful because it allows the analysis of

34 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/an-opinion-mining-approach-for-drug-reviews-in-spanish/308512](http://www.igi-global.com/chapter/an-opinion-mining-approach-for-drug-reviews-in-spanish/308512)

## Related Content

---

### Extended Adaptive Join Operator with Bind-Bloom Join for Federated SPARQL Queries

Damla Oguz, Shaoyi Yin, Belgin Ergenç, Abdelkader Hameurlainand Oguz Dikenelli (2017). *International Journal of Data Warehousing and Mining* (pp. 47-72).

[www.irma-international.org/article/extended-adaptive-join-operator-with-bind-bloom-join-for-federated-sparql-queries/185658](http://www.irma-international.org/article/extended-adaptive-join-operator-with-bind-bloom-join-for-federated-sparql-queries/185658)

### Applied Analytical "Distant Reading" using NVivo 11 Plus™

Shalin Hai-Jew (2017). *Social Media Data Extraction and Content Analysis* (pp. 159-201).

[www.irma-international.org/chapter/applied-analytical-distant-reading-using-nvivo-11-plus/161964](http://www.irma-international.org/chapter/applied-analytical-distant-reading-using-nvivo-11-plus/161964)

### Big Data Paradigm for Healthcare Sector

Jyotsna Talreja Wassan (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 570-587).

[www.irma-international.org/chapter/big-data-paradigm-for-healthcare-sector/150182](http://www.irma-international.org/chapter/big-data-paradigm-for-healthcare-sector/150182)

### Discovering Surprising Instances of Simpson's Paradox in Hierarchical Multidimensional Data

Carem C. Fabrisand Alex A. Freitas (2006). *International Journal of Data Warehousing and Mining* (pp. 27-49).

[www.irma-international.org/article/discovering-surprising-instances-simpson-paradox/1762](http://www.irma-international.org/article/discovering-surprising-instances-simpson-paradox/1762)

### Deep Learning Approach Towards Unstructured Text Data Utilization: Development, Opportunities, and Challenges

Shikha Jain, Shubham Jainand Ajit Kumar Jain (2021). *New Opportunities for Sentiment Analysis and Information Processing* (pp. 29-49).

[www.irma-international.org/chapter/deep-learning-approach-towards-unstructured-text-data-utilization/286903](http://www.irma-international.org/chapter/deep-learning-approach-towards-unstructured-text-data-utilization/286903)