

Sentiment Analysis tool for Pharmaceutical Industry & Healthcare

¹Hanane Grissette, ¹EL Habib Nfaoui, ²Adil Bahir

¹LIAN Laboratory, Sidi Mohamed ben Abdellah University
Fez Morocco.

²IT Head/Business Information Manager,
Novartis pharma, Casablanca Morocco.

Hanane.grissette@usmba.ac.ma; Elhabib.nfaoui@usmba.ac.ma; adil.bahir@novartis.com

ABSTRACT

Sentiment analysis (SA) is broadly used to analyze people's opinions about a product or an event to identify breakpoints in public opinion. Particularly, pharmaceutical companies use SA to ensure they gain a competitive edge through better understanding of patients' experiences allowing for more personalization and high responsiveness to consumers on social media. Patients self-reports on social media, frequently capture varied elements ranging from medical issues, product accessibility issues to potential side effects. The exploitation of such overwhelming unstructured data on social media through SA is of critical importance, however, the general-purpose sentiment analysis tools are not adapted and do not incorporate the specific lexicon used in the life sciences/pharma context which reduces the ability of such tools to accurately detect the meanings of the sentiments expressed towards treatments/scientific studies and pharma companies at large. Indeed, those tools involve generalized dictionaries, techniques and methods that face several challenges to detecting the suitable sentiment polarity. In this paper, we develop a dedicated research tool that extract in real time the sentiments and emotions conveyed by users on social media regarding pharmaceutical industries. Our focus is to improve the analyze of opinion about a target in pharmaceutical industries based on hybrid approaches used in single unified system throughout sentiment analysis process to seek the relevant sentiment polarity, which is tailored specifically to detect positive, negative or neutral opinions of patients and consumer health. This paper proposes a first architecture validated by a real and industry case.

Keywords: Pharmaceutical industry, Sentiment Analysis, real-time analytics, healthcare, public mood, Social media.

1 Introduction

Sentiment analysis (SA) has received great attention recently, due to the huge amount of user-generated information on social Media, which deals with the computational treatment of opinion, sentiment and other components. Mining social media presents a direct response to the surge of interest in understanding the conversations, identifying the relevant content and acting appropriately on it to identify new opportunities, deal with new business solutions and manage reputations. The Pharmaceutical Industry is no exception and SA is changing the game for Pharma companies in terms of

understanding the sentiments and the conversation around their products and is becoming a key pillar in their digital strategy. TABLE I illustrates key highlights and issues faced by patients or HCPs that emerged through posts on Facebook & Twitter.

Several recent publications assume to the richness of information to be found in patient self-reports of their problems in social media, and the useful reports are enhanced, thus absolutely aiding pharmaceutical industries to get patient voice and capture unpreventable issues.

Although there are several studies and tools of SA, few tools in the field of pharmaceutical industries use SA to analyze opinion considering the person himself or other entities as target of analysis[1]. Thus, the performance of SA tools in this field is mainly related with polarity classification task, when bad scoring of terms can effectively affect the global sentiment polarity, that is occurs when a piece of text stating an opinion on a single issue that can be classified as one of two opposing sentiments regarding to a specific context. However, according to specific industry, the target data to be analyzed can have a different distribution than the data-driven dictionaries offered. Indeed, polarity classification must identify professional and colloquial expressions and terms based on both dictionaries and machine learning techniques to improve the quality of detecting the suitable sentiment polarity, making the product evaluations more credible. Even though, the accurate meaning of expressions based on context which it appears that has great importance in determining the relevant polarity of sentiment.

Recognizing the importance of mining social media using SA techniques for pharmaceutical industries, we are motivated to develop a dedicated research tool that improve the detection of positive or negative people's opinions regarding pharmaceutical industries in real time, by inspecting posts of authors, medical entities and posts about events. This tool involves several APIs, that's based on statistical learning techniques and lexicon based approaches.

Table 1. Examples of Patients Posts On Social Media

PostID	Post content	Type of post	Report source
844476 028055 703552	Woke up too late to take my Concerta [Drug] yesterday so I'm much just slept through the entire day and only woke up an ho	Patient Experience	Cancer patient
844247 270870 716418	Where do people get Concerta [drug] ? hey i have inattentive adhd and im on meds (concerta) but im having a problem about...	Drug/Product Accessibility issue	Patient
843824 219754 381312	evr to be clear...I'm on Concerta [drug] and in the official name it has ' Methylphenidate '. It's a legal form of it....	Patient Experience	Hearsay
844536 299726 032896	Novartis [Pharmaceutical company] heart-failure med #Entresto cuts Hb A1c ,insulin starts in diabetics FiercePharma	HCP feedback	Diabetes
844525 854600 290304	Novartis [Pharmaceutical company] heart drug portfolio hit by failed serelaxin[investigational drug] study – Reuters.	Event	Fitness news magazine

The remainder of this paper is organized as follows, Section 2 presents and summarizes related works. Section 3 describes SA process. Section 4 describes our proposed tool. Section 5 presents results, discussions, limits and perspective as research questions in SA field. Section 6 concludes this paper.

2 Related Work & Applications

SA become interested part of real-life applications in various areas like business intelligence, market prediction, Recommender System, Marketing intelligence, attrition prediction.

Some works were done about SA, especially comparing the tools proposed. The work presented in [2] makes the comparison of nine SA tools: AchemyAPI, Lymbix, MLAnalyzer, Repustate, Semantria, Sentigem, Skytle, Textalytics and Textprocessing. To calculate the accuracy of each tool, texts of different sources were collected (news, comments and tweets). The tools with the greatest accuracy were Textalytics (75%), Skytle (73%) and Semantria (68%).

In other work, [3], SA tools were also compared. Twenty tools

were chosen: fifteen stand-alone SA tools (SentiStrength, Chatterbox, Sentiment140, Textalytics, Intridea, AiApplied, ViralHeat, Lymbix, SentimentAnalyzer, TextProcessing, Semantria, uClassify, MLAnalyzer, Repustate and a last one referred to as Anonymous by the authors) and five workbench tools (BPEF, Lightside, FRN, EWGA, RapidMiner). The texts used to analyze the tools were tweets related to the themes: telecommunications, pharmaceutical, security, technology and consumer products at retail. Among the stand-alone tools, the one with the greatest average accuracy (67%) was the SentiStrength. Among the workbench tools, BPEF presented best average accuracy (71%). Another study is Ref. [4]. This tested AlchemyAPI, OpenAmplify and Texterra on public text. The last one was a new tool proposed by the article and presented accuracy of 79%, higher than the OpenAmplify (57%) and AlchemyAPI (42%).

Another Study is Ref. [5]. Were conducted on existing SA methods, they measure the percentage of agreement for each method from the messages classified for each pair of methods, for what fraction of these messages they agree. They find that some methods have a high degree of overlap as in the case of LIWC and PANAS-t (80%), while others have very low overlap such as SenticNet and PANAS-t (30.77%). PANAS-t and Happiness Index had no intersection. In summary, the above results indicate that existing tools vary widely in terms of agreement about the predicted polarity, with scores ranging from 33% to 80%. This implies that the same social media text, when analyzed with different sentiment tools, could be interpreted very differently. In particular, for those tools that have lower than 50% agreement, the polarity will even change accordingly to pharmaceutical industries (e.g., from positive to negative, or vice versa). Nevertheless, several startup that measure opinion on real data have emerged and a number of innovative products related to this theme have been developed. There are

multiple methods for measuring sentiments, including lexical-based and supervised machine learning methods [6].

Despite the vast interest on the theme and wide popularity of some methods, it is unclear which one is better for identifying the polarity for posts. Indeed, pharmaceutical, medical professionals and health vocabularies are often observed that are different from those of public [7]. To interpret medical terms in online patient discussions, Consumer Health Vocabulary (CHV), a lexicon linking UMLS standard medical terms to patients' colloquial language [8], is adopted in few recent studies. We summarize in TABLE II the recent work of SA in health and pharmaceutical field. We have selected three criterions: methods, approaches and data source.

3 Sentiment Analysis Process

In this section, we describe the broad dimensions and the process of SA. As defined in [19], sentiment analysis is a computational study of opinions, sentiments, Subjectivity, emotions, and attitude expressed in texts towards an entity. Thus, SA is a multi-task of detecting, extracting and classifying opinions, sentiments and attitudes concerning different topics, as expressed in textual input. In Fig 1. presents seven broad dimensions of SA.

Table 2: Works Summary of Related Work In Pharmaceutical & Healthcare

Prior Study	Description	Data Source/Methods/approaches
Benton et al. [9] (published in 2011)	They proposed advanced techniques for analyzing health social media content.	They adopted three breast cancer forums as test bed. Using Lexicons: CHV, FAERS for medical entity Extraction approaches.
Mao et al. [10] (published in 2013)	In this work, the proposed architecture Help to understand drug switching and discontinuation behaviors to understand patient reported adverse drug events.	They developed their test bed on 12 breast cancer forums. Using Lexicons: CHV, FAERS for medical entity Extraction approaches.
Sarker et al. [11] (published in 2015)	The proposed method improves detection of sentences with ADE mentions, they developed a text classification method for adverse drug reaction detection on clinical reports.	Data source is obtained from clinical reports, Twitter, and daily strength. Using Lexicons: UMLS, WordNet, MedEffect, SIDER, and COSTART.
Ramon et al. [1] (published in 2016)	This study present a Sentiment Analysis tool, named SentiHealth-Cancer (SHC-pt) that improves the detection of emotional state of patients in Brazilian online cancer communities, by inspecting their posts written in Portuguese language.	The SHC-pt tool is based on the method called SentiHealth proposed in ther paper [19]. Moreover, as a second alternative is to analyze the texts in Portugues.
Segura-Bedmar et al. [12] (published in 2015)	They developed a distant supervision approach to extract adverse drug events in Spanish	Data is collected from Forums and blogs, they used GATE pipeline for medical entity recognition and distant supervision with shallow linguistic kernel for adverse drug event extraction.
Chee et al. [13] (published in 2011)	They developed ensemble classifiers with SVM and nave Bayes to classify risky drugs and safe drugs based upon online discussions	Data is collected from Health forums in Yahoo! Groups; SVM and Naive Bayes are used for classification tasks; Lexicons: UMLS, MedEffect, SIDER for medical entity recognition approaches.
Wu et al. [14] (published in 2016)	They developed a discriminant classifier with Ricchio method and a generative model to determine whether the side effect is relevant to the drug	Data Source is collected from online discussion; Rocchio method is used for classification task.
Nikfarjam et Gonzalez [15] (published in 2011)	They proposed a method to capture the underlying syntactic and semantic patterns from the Daily Strength reviews.	Reviews are collected from social media e.g. "Daily strength"; association rule mining is used for entity recognition task & approaches to generate patterns for adverse event recognition. This study benefited from a large, manually annotated corpus (1,200 records) and an algorithm that can detect expressions not included in a lexicon.

<p>K. Portier et al [16] (published in 2013)</p>	<p>The objective of this study is understanding topics and sentiment, detect pessimistic emotional state. And also detect changes in a person mood as consequence of his interactions with other patients in a community. That is used to give emotional support to patients.</p>	<p>SA techniques applied on posts in cancer online communities.</p>
<p>Liu et all .[17] (published in 2011)</p>	<p>They are classifying Yahoo Answers posts into patients' posts or medical professionals' posts 82.0% in f-measure for identifying drug user from tweets.</p>	<p>They used test beds such as Twitter, hashtags and URLs are adopted in the feature set. They used sentiment related words from LIWC as features in classification. In terms of learning methods, the most commonly employed learning algorithms are Support Vector Machines (SVM) and naive Bayes.</p>
<p>Abdul-Mageed et al. [18](published in 2014)</p>	<p>This study Performed subjectivity and sentiment analysis (SSA) of social media for a morphologically-rich language.</p>	<p>They Worked on 2798 chat turns, 3015 Arabic tweets, 3008 sentences from 30 modern standard Arabic Wikipedia Talk pages, and 3097 web forum sentences.</p>

3.1 Subjectivity Classification

Subjectivity classification is the property associated with words and word sense that deals with the detection of “private states”— a term that encloses sentiment, opinions, emotions, evaluations, beliefs and speculations. To deal with opinion word (features) and product aspect extraction considered named entity recognition (NER) system for streaming micro-text. It was based on different contextual associations like semantic association, temporal association, and social association between micro-text clusters.

3.2 Sentiment Classification:

Sentiment classification is the determination of orientation of sentiment of given text in two or more classes.

3.3 Polarity determination

Sentiment classification is concerned with determining polarity of a sentence, whether a sentence is expressing positive, negative or neutral sentiment towards the subject. Hence, Sentiment classification is also termed as polarity determination.

3.4 Opinion feature extraction and product aspects extraction

On social networking, if someone is linked to/followed/liked a social networking group or community or public figure, it implies that the person has either positive or negative feelings

Towards that entity. To perform finer grained SA regarding to aspect of product and opinion word as features, we need to identify feature's opinion and its Sentiment score of different aspect that will affect global sentiment strength.

Otherwise, Sentiment analysis is not a single problem, various steps are needed to perform Sentiment Analysis from given texts, are presented in Fig 2.



Figure 1: the seven broad dimensions of SA

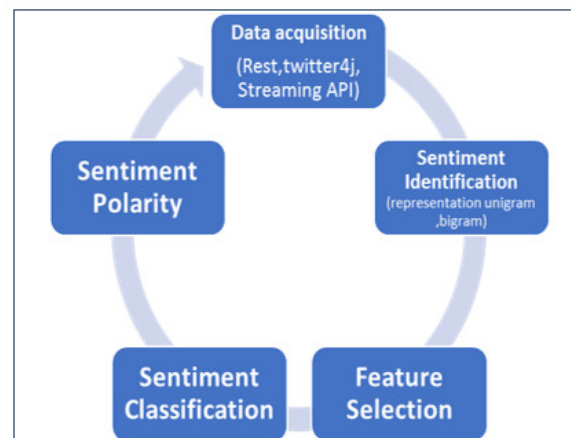


Figure 2: Sentiment analysis steps

SA techniques can be roughly divided into machine learning approach, lexicon based approach and hybrid approach. The Machine Learning Approach (ML) applies the famous ML algorithms and uses linguistic features. The Lexicon-based Approach relies on a sentiment lexicon (a collection of known and precompiled sentiment terms), that's divided into dictionary-based approach and corpus-based approach which use statistical or semantic methods to find sentiment polarity. The hybrid Approach combines both approaches and is very common with sentiment lexicons playing a key role in most methods.

4 Proposed Architecture & Process

The objective of this work is to present a SA tool, that improves the detection of opinions regarding pharmaceutical industry in real time and analyse of emotional state of peoples talking about, to get up to the minute information on social media. That is explained in detail below Fig 3.

4.1 Pharmaceutical industry data collection:

There are several social networks that people can use to interact with each other. The most widely used of these are Facebook and twitter. In January 2017, the network contained 1.86 billion monthly active Facebook users, where an active user is one who has realized log-in in the last 28 days and also has at least one friend added to his network. From a recent study, 54% of patients are very comfortable with their

providers seeking advice from social networks to better treat their conditions. In other study, more than 40% of consumers say that information found via social media affects the way they deal with their health.

Another advantage of using Facebook and twitter are these they have a lot of patients and they are very active considering the content disseminated in them. For that, we developed an automated crawler program to collect data from Facebook & Twitter. By inspecting opinions of pharmaceutical industries, our collector program is based-keyword to seek opinions about specific products, consumer health and events in the pharma world. Collected data includes post ID, URL, post content, post author's ID and post date.

As our study focuses on sentence level information extraction and processing, we handle our source data obtained from crawler program in table and parse post content to be an input of next step of analysis.

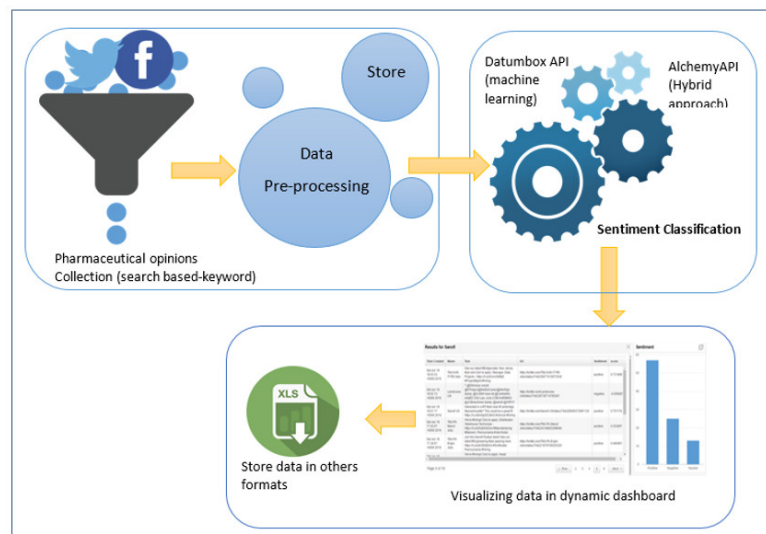


Figure 3: Major Components of proposed architecture

4.2 Sentiment Classification:

It is a challenging task to classify posts from noisy patient-generated content and professional, according to medical and pharmaceutical expressions. Lexicon based approach was the best performing medical entity recognition system in prior studies [20]. Our tool incorporates the statistical learning methods and machine learning approaches also known as hybrid approaches to identify the suitable sentiment polarity. Indeed, proposed tool involves methods from Alchemy API & Datumbox. When AlchemyAPI incorporates statistical analysis, techniques based on machine learning algorithms, and linguistic techniques from natural language processing into a single unified system. It support to English and other languages and can be used by applications programmed with java, C/C++, C#, Perl, Java, PHP, Python, JavaScript, Ruby and can also be used by android applications [51]. A free version is available that allows to perform up to a thousand daily transactions and five concurrent request. In addition, Datumbox API is a machine learning Framework presents a several Text Classifiers such as the Naive Bayes, the SoftMax Regression and the Max Entropy used to classify Sentiment based on statistics methods.

Through parameters, to perform SA on document level (reviews and posts are considered as document). We present the suitable sentiment polarity of each document on dynamic dashboard. The results of the analysis may be output in excel formats, to deal better with other purpose.

5 Case Study

5.1 Novartis international company in Maghreb

This work is part of a project evaluated and approved by industry case, one of the global leader in the pharmaceutical industry, Novartis International AG in Morocco country that is a Swiss multinational pharmaceutical company. With the growing availability and popularity of opinion-rich resources such as social media. Table III presents a summary of search volumes for given keyword “Novartis” to assess the potential search engine traffic for Novartis brand on the web.

Table 3: Daily Volume of Search Keyword Novartis from Wordtracker

keyword	Volume
novartis	1,428
Novartis Emcurevysov	108
Novartis Emcure partnership	106
Novartis and vysov	104
Novartis usv partnership	86
Novartis ophthalmology	65
Novartis rebate	56
Novartis osteoporosis	41

Otherwise, our statistics performed on Facebook and twitter presented in Fig 4 that both show the strong presence represented by sum of Facebook fans and Twitter followers in different pages. Other statistics track user interactions on dedicated pages of pharmaceutical industry on Facebook and Twitter respectively that are presented in Fig 5. And Fig 6. Nevertheless, Novartis pharma industry have very motivated to poll these microblogs to get insight from people’s experiences and related-drug problems of their product and performed events... etc.

There are few works proposing context-domain SA solutions aiming at improving accuracy of the classification of opinions in the pharmaceutical context that involves driven-domain methods and approaches to get right sentiment context. Based on the research gaps, to adapt generalized-purpose SA tools at local context, that is identified by many industries cases. For that, we are motivated to adapt our tool at Novartis pharma context, the final objective is to get insights of minute information about pharmaceutical industries on social media, whether its Novartis pharma, its products, its event or its competitors in the world.

5.2 Validation and Experimental Result

In this section, we present the results obtained by applying tool on the different collections of data collected from Facebook & twitter. The conducted experiments consider the language of posts are written, the most of analyzed texts are in English language. (For example Fig 7).



Figure 7: tweet of Ritaline [Drug] detected in English language.

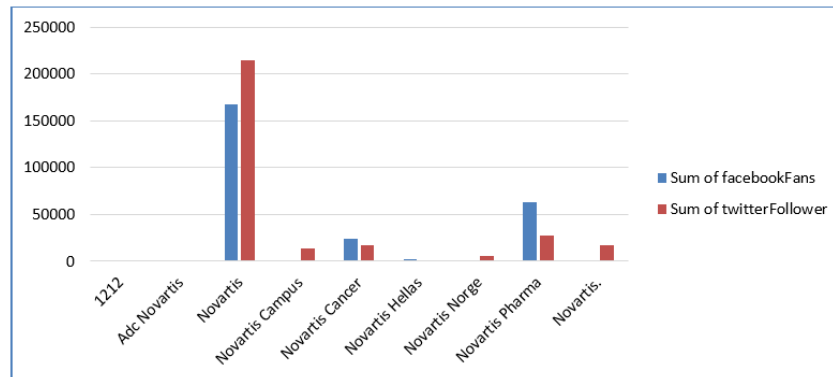


Figure 4: sum of Facebook fans and Twitter followers in different pages

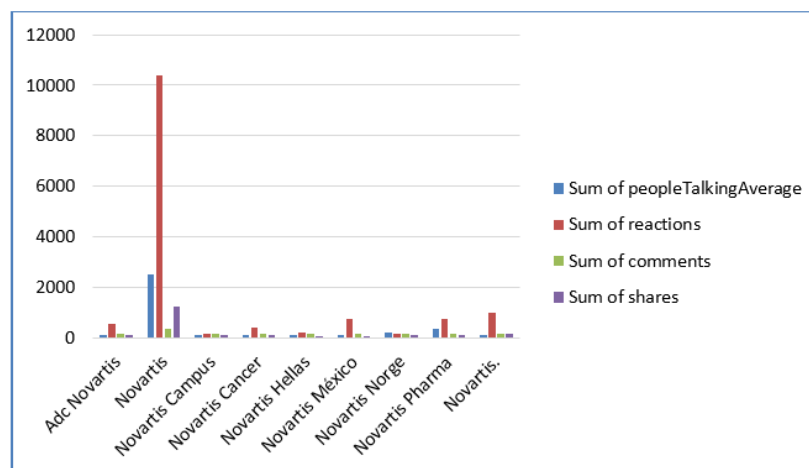


Figure 5: interaction rate by weekday on all dedicated Novartis pages

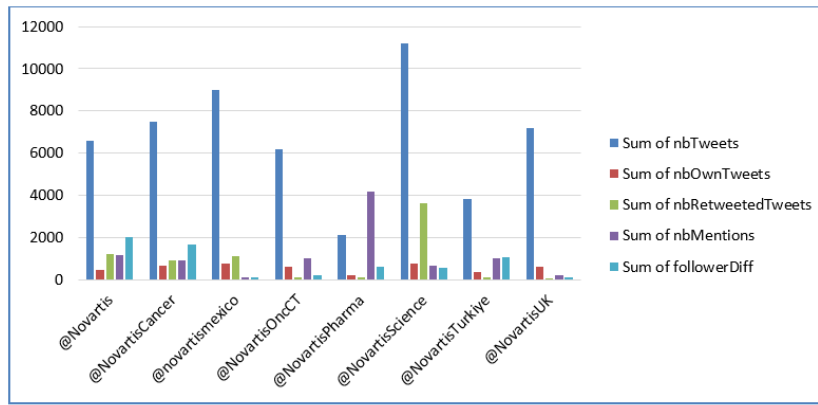


Figure 6: average Followers on dedicated Novartis pages in twitter and its interactions

Facebook and twitter are the most popular platform used by patients and professionals to share their experiences. For that, we conduct our experiments to extracting people-report on Novartis products & events, our crawler program randomly selected 100 posts in real time based on keyword. The result presented in TABLE V summarizes the results of some experiments performed on Facebook pages in real time. In addition. As shown in Fig 8. The 100 post collected considering “Novartis” opinions (the crawler program must collected all tweet contains it), 30 post are positive, 55 are neutral and 15 are negative. Others results presented bellow shows there is a mix of sentiment in the collected post.

Table V: Results Obtained of the Experiments Facebook

POSTS DATE	KEYWORD USED	POSITIVE	NEGATIVE	NEUTRAL
FRI ABRIL 18 13:42:01	Novartis	30	15	55
FRI ABRIL 18 13:48:01	Novartis pharma	32	12	65
FRI ABRIL 18 13:50:01	Concerta [dug]	7	5	88
FRI ABRIL 18 14:42:01	Novartis+ Sanofi [pharmaceutical company]	38	22	40

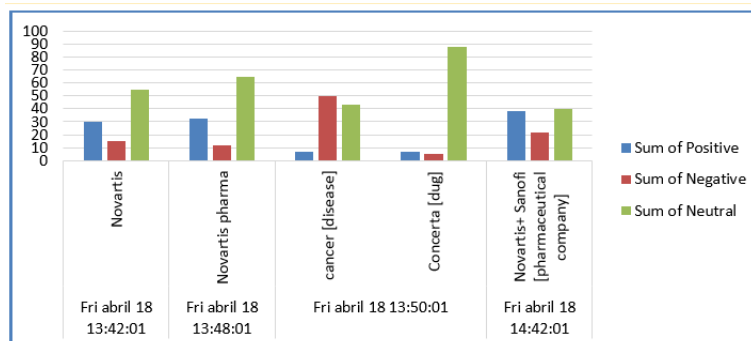


Figure 8: average of sentiment polarity of opinions in Facebook pages

Table 4: Results Obtained of the Experiments on Twitter

POSTS DATE	KEYWORD USED	POSITIVE	NEGATIVE	NEUTRAL
SAT JUN 18 13:38:01	Novartis	63	17	20
FRI ABRIL 18 13:48:01	Novartis pharma	35	50	15
SAT JUN 18 13:50:01	Alcon[Novartis Filial]	42	28	30
SAT JUN 18 15:42:01	Sanofi[American pharmaceutical company]	57	25	18
SAT JUN 18 15:43:01	Pfizer[American pharmaceutical company]	58	25	17
SAT JUN 18 14:45:01	GSK[British pharmaceutical company]	50	23	27
SAT JUN 18 10:24:01	Concerta [dug]	19	60	21
SAT JUN 18 10:22:01	Exelon[dug]	42	17	41
SAT JUN 19 10:32:01	Jakavi[dug]	5	1	1
SAT JUN 19 10:42:01	Ritalin[dug]	40	45	15

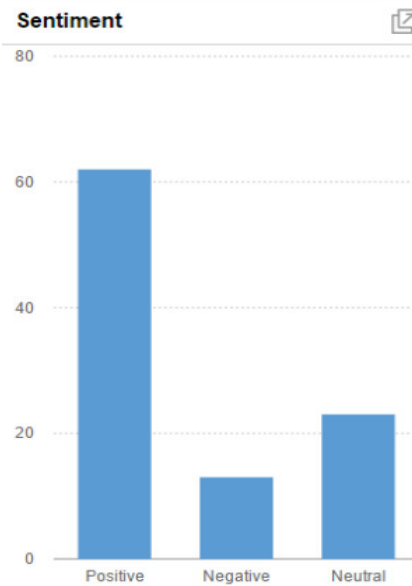


Figure 9: Screenshot of dynamic dashboard global polarity sentiment obtained by tool

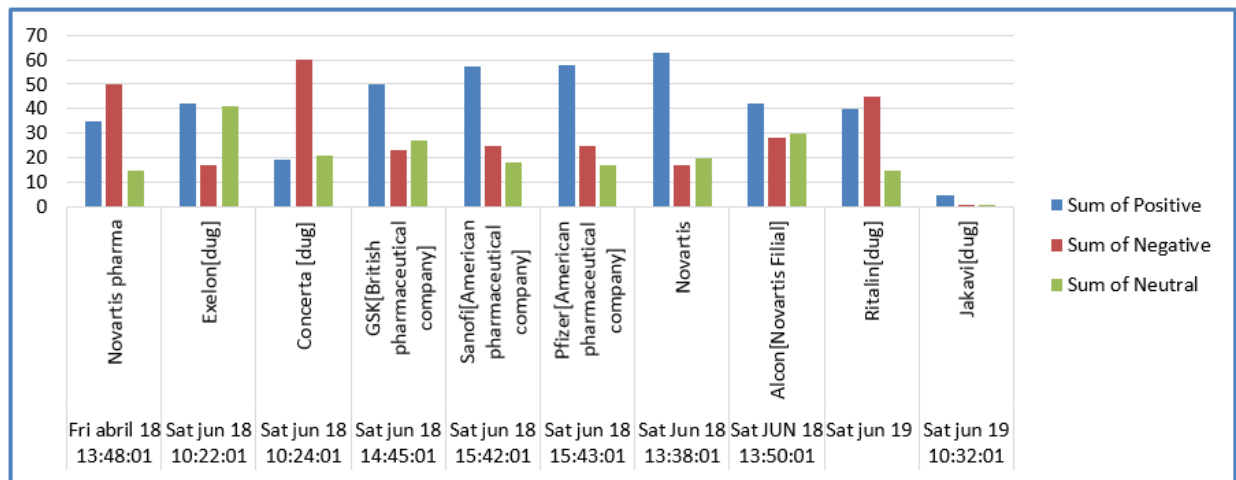


Figure 10: summarize some people’s opinions on twitter at 18/19 June in 2016

Regarding the context of pharmaceutical & health, many people usually choose twitter to share their problem or to search information about drug-treatment, for that we conduct another experiments to hear people’s voice on twitter, TABLE IV present some results obtained at 18 in June 2016. From Fig 10. It’s clear that on twitter people are more interactive considering twitter as credible microblogs to share and change health and treatment information.

Fig 11. Showeda screenshot of data collected by our crawled program and their classification obtained in our prototype, which is based on different keyword to seek opinions about them. Accompanied with a dynamic dashboard to visualize results that show the global sentiment polarity of opinion Fig 9.

Time Created	Name	Text	Url	Sentiment	sc
Sat Jun 18 13:42:21 +0000 2016	Patrick O’Sullivan	@Novartis @MedsforMalaria @gatesfoundation You lost me at "Gates Foundation". #ThanksButNoThanks RT @Novartis: We’re not giving up in the fight against antibiotic resistant superbugs: https://t.co/ZL9RAFrAiu #ASMMicrobe2016 https://t.co...	https://twitter.com/PatrickO’Sullivan/status/744163343146049536	negative	
Sat Jun 18 13:40:02 +0000 2016	Miran Šinigoj	RT @Cannes_Lions: Some words of wisdom from @Novartis from the Inspiration Stage #LionsHealth #CannesLions https://t.co/AtopdWk...	https://twitter.com/MiranŠinigoj/status/744162760217477120	neutral	
Sat Jun 18 13:38:15 +0000 2016	Caroline Arnaud	RT @Cannes_Lions: Some words of wisdom from @Novartis from the Inspiration Stage #LionsHealth #CannesLions https://t.co/AtopdWk...	https://twitter.com/CarolineArnaud/status/744162312249024512	positive	

Page 1 of 10

Figure 11: Screenshot of tweets dataobtained by tool and their classification obtained.

The accuracy is mainly based on the quality of the annotation, word’s score, and recognition entity. In order to better adopt this tools in a specific domain, usually the training process will take a long time to summarize their customized entity of each domain and term at all.

Besides, the performance of algorithms and techniques of sentiment analysis is usually not good when oriented to a local context or domain such as score of entities changed in some context. Also, Such Brand

or Industry has its knowledge base and its concepts, especially when we spoke about Pharmaceutical industry. There is word's polarity change from context which it appears to another. Therefore, the meaning and context's terms used in posts about product's Novartis can affect global polarity. For example, the tweet at Fig12 below has been classified as negative tweet by AlchemyAPI, but tweet present a very important result offered by Novartis study, so it's a positive tweet for Novartis. Even if it talks about patients with the most serious form of childhood arthritis.

We notice that the same gaps exist in another domain. For example, in [20] an automobile organization have a same gap of analysis, that is concerned only about recognizing the sentiment information received for automobiles only. Therefore, a list of Domain Dedicated PolarWords (DDPW), proved as the best lexical resource for domain oriented sentiment analysis.

The research work presented in this paper allows to analyze the sentiment expressed towards pharmaceutical products according to local context. Therefore, as a future extension, we will propose a dynamic method to recognize semantic meaning according to local context, regarding to medical entity, persons, and drug's name, adverse drug event (ADE) and others features. In this context, some research question is remaining to resolve:

- What are the requirements and methodologies for developing extensible and scalable SA Model in real time?
- How statistical learning techniques can allows to create dynamic dictionaries regarding pharmaceutical and medical entities?
- What are considerations should be taken into account for improve the quality of SA according to local context?

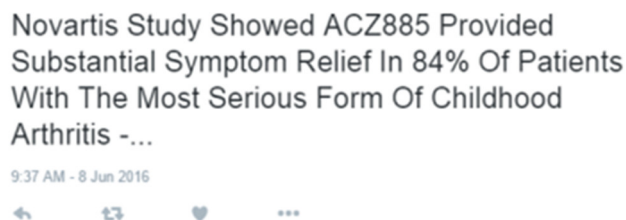


Figure 12. Tweet about patients with the most serious form of childhood arthritis.

ACKNOWLEDGMENT

Thanks to Novartis Pharma Company in Morocco for partnering and supporting the present research.

6 Conclusion

The ability to extract insights from social data is a practice that is being widely adopted by organizations across the world.

In the proposed tool, we classified sentiments and emotions conveyed by users in real time as positive, negative and neutral, regarding the pharmaceutical industry. Many experiments were done towards different Novartis's product and its competitors. It would be interesting to extend this work to consider more features that improve the quality of classification of sentiments and recognition of entities. This makes possible the SA establish other polarities that show more clearly personal emotions.

REFERENCES

- [1] Ramon et al. "SentiHealth-Cancer: A sentiment analysis tool to help detecting mood of patients in online social networks", Published by Elsevier Ireland, <http://dx.doi.org/10.1016/j.ijmedinf.2015.09.007>.
- [2] Cieliebak et al. , "Potential and limitations of commercial sentiment detection tools", Proceedings of the First International Workshop on Emotion and Sentiment in Social and Expressive Media: approaches and perspectives from AI (ESSEM 2013), A Workshop of the XIII International Conference of the Italian Association for Artificial Intelligence (AI*IA 2013), Turin, Italy, December 3, 2013, 2013, pp. 47–58 <http://ceur-ws.org/Vol-1096/paper4.pdf>.
- [3] Abbasi et al. "Benchmarking twitter sentiment analysis tools", Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014), Reykjavik, Iceland, May 26–31, 2014, 2014, pp. 823–829 <http://www.lrec.org/proceedings/lrec2014/summaries/483.html>.
- [4] Turdakov et al. "A framework for text analysis", Program. Comput. Software 40 (5) (2014) 288–295, <http://dx.doi.org/10.1134/s0361768814050090>.
- [5] Pollyanna et al. "Comparing and Combining Sentiment Analysis Methods", not published.
- [6] Montoyo et al., "Subjectivity and sentiment analysis: an overview of the current state of the area and envisaged developments", *Decis Support Syst.* 53 (2012) 675–679.
- [7] KANG et al. "Review-based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach", *Expert Syst. Appl.* (2013), [HTTP://DX.DOI.ORG/10.1016/J.ESWA.2013.07.101](http://dx.doi.org/10.1016/j.eswa.2013.07.101)
- [8] KUHN et al., "A side effect resource to capture phenotypic effects of drugs", *Mol. Syst. Biol.* 6 (1) (2010) 343.
- [9] BENTON et al., "Identifying potential adverse effects using the web: a new approach to medical hypothesis generation", *J. Biomed. Inform.* 44(6) (2011) 989–996
- [10] MAO et al. "Online discussion of drug side effects and discontinuation among breast cancer survivors", *Pharmacoepidemiol. Drug Saf.* 22 (3) (2013) 256–262.
- [11] SARKER et al. "Portable automatic text classification for adverse drug reaction detection via multi-corpus training", 53 (2015) 196–207.
- [12] SEGURA-BEDMAR et al., "Exploring Spanish health social media for detecting drug effects", *BMC Med. Inform. Decis Making* 15 (Suppl. 2) (2015) S6
- [13] CHEE et al., "Predicting adverse drug events from personal health messages", *AMIA Annual Symposium Proceedings*, 2011, American Medical Informatics Association, 2011, p. 217.
- [14] WU et al., "Exploiting online discussions to discover unrecognized drug side effects, *Methods Inform. Med.* 52 (2) (2013) 152–159.
- [15] Nikfarjam, A., & Gonzalez, G. H. (2011). Pattern mining for extraction of mentions of adverse drug reactions from user comments. In *AMIA Annual Symposium Proceedings* (Vol. 2011, p. 1019). American Medical Informatics Association. K. PORTIER, G.E. GREER, L. ROKACH, N. OFEK, Y. WANG,

- [16] P. BIYANI, M. YU, S. BANERJEE, K. ZHAO, P. MITRA, J. YEN, "Understanding topics and sentiment in an online cancer survivor community", *J. NATL. CANCER INST.—MONOGR.* 47(2013) 195–198.
- [17] F. Liu, L.D. Antieau, H. Yu, Toward automated consumer question answering: automatically separating consumer questions from professional questions in the healthcare domain, *J. Biomed. Inform.* 44 (6) (2011) 1032–1038
- [18] M. ABDUL-MAGEED, M. DIAB, S. KÜBLER, SAMAR, "Subjectivity and sentiment analysis for arabic social media", *COMPUT. SPEECH LANG.* 28 (2014) 20–37
- [19] Asghar MZ1, Ahmad S2, Qasim M1, Zahra SR1, Kundi FM1. SentiHealth: creating health-related sentiment lexicon using hybrid approach. 10.1186/s40064-016-2809-x 2016.
- [20] R. Leaman, L. Wojtulewicz, R. Sullivan, A. Skariah, J. Yang, G. Gonzalez, Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks, in: *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*, Association for Computational Linguistics, 2010, pp. 117–125.