

**Faculty of Informatics and Computer Science**

*Computer Science*

**Mining Post-Marketing Adverse Drug Reactions on Social Media Using Text Processing and Machine Learning**

By: Ahmed Hossam ElSabbagh

Supervised By

Dr. Nahla Barakat

**November 2018**

Abstract

Adverse Drug Reactions/Events (ADR) are a harmful or unpleasant reaction caused by the use of medicinal product [1], these effects are studied through the science of Pharmacovigilance, the science and activities relating to the detection, assessment, understanding and prevention of these effects [2] [3]. The ADR caused by drugs after their release is considered a major health hazard, causing a large number of death, it is estimated that more then 6% of patients are hospitalized as a cause of serious ADRs, which is translated into more than 2 million patients, 0.32% of whom are fatalities or more than 100 thousands deaths in the US alone [4].

This project aims to research and find a good method to implement an Adverse Drug Reaction (ADR) detection program using machine learning based methods, with dataset collected from social media and/or medical forums.

Introduction

## Overview

Adverse Drug Reactions/Events (ADR) are a harmful or unpleasant reaction caused by the use of medicinal product [1], these effects are studied through the science of Pharmacovigilance, the science and activities relating to the detection, assessment, understanding and prevention of these effects [2] [3]. The ADR caused by drugs after their release is considered a major health hazard, causing a large number of death, it is estimated that more then 6% of patients are hospitalized as a cause of serious ADRs, which is translated into more than 2 million patients, 0.32% of whom are fatalities or more than 100 thousands deaths in the US alone [4].

There are many steps taken Through medical trials to find the ADRs, however some people may not have the same side effects as others and therefore not all ADRs that affect all different patients appear until phase IV trials -postmarketing trials- especially since most medical trials are concentrated on certain demographics, therefore are many uncertainties of the effect of the drugs on any given population [5] [6], and despite the many mediums offered by the FDA to report ADR (FAERS, MERP, MedWatch) [3], 90% of the ADR are in fact under-reported [7].

However, a new field of pharmacovigilance via social media has been introduced in recent years, as there are many disease support networks (DailyStrength and MedHelp), patient forums (AskAPatient) and miniblogs(Twitter) [8], where patients are involved in sharing their experiences with certain drugs, with many of them and their caregivers actively read said experiences [3]. Many data mining techniques were adapted to extract potential ADRs of drugs by through text mining and machine learning, allowing many researchers not only a new way research ADRs, but also help them find new ADRs they had not known to have previously existed in some drugs or find earlier occurances then previously reported [3] [9].

## [Scope](http://www.cs.stir.ac.uk/~kjt/research/conformed.html) and Objectives

This project aims to research and find a good method to implement an Adverse Drug Reaction (ADR) detection program using machine learning based methods, with dataset collected from social media and/or medical forums.

# Related Work (State-of-The-Art)

## Background

There has been a lot of work related concerning ADR datamining, the most prevalent methodologies include lexicon-based pattern mining [5] [6] [9] [10] and supervised machine learning [11] [12] [13] approaches [3], however there are also rare cases of hybrid systems [14] [15], and partially supervised implementations [16]. The evaluation of these techniques was done mainly using three metrics, F-Score, Recall, and Precision [17] [18] [19].

## Literature Survey

**Traditional ADR Detection Methods**

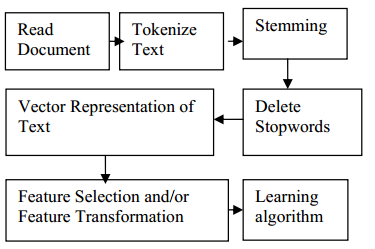
## Electronic post marketing pharmacovigilance has traditionally been applied through FDA Adverse Event Reporting System (FAERS) [20], MedWatch [21] and the Institute of Safe Medication Practices Medication Error Reporting System (MERP) [3]. These methods however are voluntary for the public and healthcare professionals, meaning that data could be missing or incomplete, and clinical narratives are limited to researchers affiliated with medical research centers [3]. However, due to the widespread use of social media, the wider knowledge and the convenience of its use, patients have been more comfortable sharing their experiences and looking for answers as shown in a survey by Pew Research Center [22]. This is why many researchers have been trying to find ways to mine social media for ADR information.

**Scope of research**

There has been two types of research, direct ADR research as surveyed by [3], which focuses on trying to discover a relation between a given drug and any ADRs, the second is drug-drug-interaction (DDI) as surveyed by [23], which focuses on the side effects of consuming several drugs in the same time. This project will focus on the first type of research.

**Finding Datasets**

The first obstacle regarding ADR research is to find the necessary data. The data used in all projects is comprised of social media posts with mentions of drug names, these posts may or may not contain mentions of ADR related to the drug name. The posts are arranged into a corpus of data which is used to train and evaluate machine learning and pattern mining approaches [3] [6] [9] [10] [11] [12] [13] [14] [16]. The social media in question are twitter, the data of which is available at Arizona State University [3] [11] [14] [24], and DailyStrength [25], MedHelp [26], PatientsLikeMe [27], Yahoo! Forums [16], Medications.Com [28] and AskAPatient [29]. Using either purpose built crawlers [5] or API [12] to get the data.

**General Text Processing Techniques:**

Generally, the following sequence of text processing techniques were used to make the data usable for learning process, with variations and different tools used to achieve it, the texts were split in the punctuations and whitespaces, creating tokens (Tokenization), stemming is applied to remove similar words or turn them to a simpler form (training, trainer, trained = train), stop words (a, and, but) are removed completely. The tokens are then turned into vectors, from which certain features are extracted and used in the learning algorithm [30].

Many researchers used lexicon for comparison, provided by UMLS Metathesaurus [31] , SIDER [32], MedEffect Canada [33] and manually allocated colloquial terms

**Lexicon-Based Techniques**

The first real attempt to create a pharmacovigilance program was in 2010 by Leaman et al. [5]. The lexicon is extracted from UMLS Metathesaurus [31] , SIDER [32], MedEffect Canada [33] and manually allocated colloquial terms [5]. It used a purpose built parallelized crawler to extract comments from DailyStrength [25], bringing about 3600 annotated comments and 450 reserved for evaluation, the information was taken for the following drugs, carbamazepine, olanzapine, trazodone, and ziprasidone [5]. The comments were annotated for adverse effects, beneficial effects, indications and other terms [5], where the following tools are using in text processing, Java (Tokenization), Snowball implementation (Stemming) [34], Jaro-Winkler measurement [35] (Similarity for misspelling) [5]. A basic lexical similarity comparison was applied, where a sliding window of tokens (size = 5) was go over the tokenized text, comparing them with lexical terms in the dictionary, pairing them in an assignment problem [5], the similarities were summed and normalized with the result by the number of tokens in lexical term [5]. The closest verbs were used to categorize the mention, where verbs like “taking” was an indication, since ADR are targeted, Indications, beneficial effects and others were filtered out [5]. 1260 adverse effects, 391 indications, 157 beneficial effects and 78 other, for a total of 1,886 annotations [5]. For the evaluation, Precision = 78.3%, recall = 69.9%, for an F-Score = 73.9%, not all known ADRs were recognized [5].

The following lexicon methods that followed use association rule, a data mining approach that tries to find statistical relation between drugs and ADRs, the goal is to find enough minimum support and confidence constraints [36]. The research by [6] [10] were among the first to apply this method.

The first research [6], which was made by the team working on [5], tried to match a comment with certain grammatical patterns in order to find the mention of ADR and detecting whether a person is actually inflicted or not [6]. To achieve that, the DailyStrength [25] dataset corpus was used, with an additional 3290 records were added for a total of 6890 records, including the original 1886 annotations from the previous research and using the same dictionary as before and the same drugs [6]. Then to solve the problem, it follows three steps: 1) Term Sequence Generation, the sequence of words in which an ADR is mentioned is stored in a file, each line in the file has the ADR replaced with ADR keyword, part of speech (POS) tagging is performed using the Stanford parser [6] [37]. Some POS were kept (like verbs) using Wordnet [6] [38]. Thus a term is created. 2) Frequent Rule Identification, Apriori tool [39] (which implements Apriori Algorithm) is used to mine association rule, where a term like “make PRP RB CC =>ADR” (which is a combination of verb make and POS tags) occurs when ADR is reported [6]. 3) Frequent Pattern Generation, where patterns based on Frequent Rule Identification are generated, with short patterns and patterns with placeholders after ADR being excluded. Precision = 70.01%, recall = 66.32% for an F-Score = 67.96%, with minimum support = 4, maximum=6, minimum number of terms per rule=4, other tunings and replacements in the input were tried, but this was by far the best result in all the attempts [6].

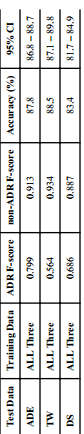
While [10] tried to detect ADRs through data in parenting sites using disproportionality techniques using the concepts discussed in the three Dutch examples methods [10] [40], the drugs used for this research were amoxicillin, paracetamol, ibuprofen, Bactrim, cetirizine, azithromycin, bacitracin, loratadine, xylometazoline [10], the data was collected from eight parenting web sites in a seven year period (2005-2012) [10], a total 1290 posts were collected, 900 annotated and 300 kept in reserve [10]. A database was created to keep drugs and posts, drugs were organized with their generic names and Anatomical Therapeutic Chemical (ATC) codes [10]. Used UMLS Metathesaurus [31] for biomedical vocabularies, and other resources such as European Agency for the Evaluation of Medical Products (EMEA), Medicines and Medical Devices Agency of Serbia (ALIMS) [41], the DrugBank Database [42], and MetaMap [10] [43], which is a lexical system that maps text to concepts in UMLS Metathesaurus. The following disproportionality measures were used to qualify drug safety: reporting ratio (RR), proportional reporting ratios (PRR) [10], reporting odds ratios (ROR) and information component (IC), each of these measures find the association between ADR and drug. The evaluation of this method resulted in precision = 75.3%, recall = 64.7%, and F-measure = 69.599% [10].

The research by [9] was unique as it tried to not only identify the ADR, but also identify when it was first mentioned [9], all using tensor decomposition to assist in the classification, having used tensors as data containers instead of matrices, greatly helping in missing data and helped identify some ADR mentions before the official FDA announcement [9]. The main advantage of this approach is that it does not require expert annotation for the data [9]. Matrix Based Technique (MBT) was compared to the proposed Tensor Based Technique (TBT). External resources are used for item extraction, mainly Consumer Health Vocabulary (CHV) for drugs and ADR lexicon [9]. Temporal factor is considered important for ADR detection and monitor association rule [9], which is why the association will be made based on a time period of a year, with value p being time window, and q being overlapping year, several combinations of both were tired to compare with TBT [9]. After the association rule mining is done with temporal analysis for each ADR, a matrix of drug x time is made where each cell is the lift measure between ADR and Drug under specific year [9]. TBT introduces using a tensor (3D matrix) between drug, ADR and time, using CANDECOMP/PARAFAC (CP) decomposition technique [9]. Tensors allow the observation of dataset in 1 year time window without overlapping, so data from 2001 and 2002 are used in the same vector, averaging the results from both years [9]. Dataset about 20 drugs were extracted from 500 threads each drug from MedHelp [26] totaling 16344 threads ranging from 1997 and 2011 [9]. Evaluation measures were not determined, but apparently several ADRs were discovered by TBT as opposed to MBT, and some of the existing ADR were discovered to have been mentioned earlier than when it was alerted by the FDA [9].

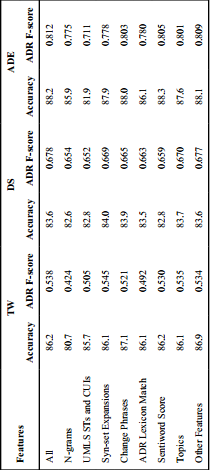
**Supervised Learning**

Next is supervised learning approaches, most ambitious ADR detection system try to track them through twitter or a mix of twitter and other sources, trying to use the large data set that can be provided through twitter [11] [12] [13]. Among the three research SVM, Naïve Bayes, and Maximum entropy were the most commonly used and evaluated.

The first to create a model that follows this approach is [44], which tried to use an ensemble of classification algorithms to classify the dataset on several stages, namely Support Vector Machines (SVM) with RFB kernel and Naïve Bayes (NB), using messages in Yahoo! Forums as dataset (exact size not determined), it used two feature sets: general vocabulary, and meta-features with specialized lexicons from MedDRA [44] [45].

The research made in [11] tried to research the ability to use multi-source dataset corpus, mainly three, twitter with data provided from Arizona State University [24], using 10822 tweets. 1082 of which were annotated by experts using the Inter Annotator Agreement (IAA) using Cohen’s Kappa [11] [46]. DailyStrength [25] with 10617 posts were used (23.7 % contain ADR mentions). And the ADE corpus which contains 23516 phrases with 29% containing ADR mentions, the corpus is not social media but it was used as a proof of concept that the system introduced can analyse data from any source [11]. Three supervised classifiers were made to test the data, Naïve Bayes (NB), Support Vector Machines (SVM), and Maximum Entropy (ME). To perform pre-processing, Porter stemmer was used using implementation by NLTK toolkit [11] [47], POS tagging was done using Stanford parser [37] and twitter parser [48]. UMLS concepts were identified using MetaMap [11] [43]. Synonymous terms were identified with WordNet [11] [38]. The features were by how often a change happens, so if an ADR is reduced, a good change happens and vice versa, so the feature set is built by identifying these changes using a sliding window on the phrase, the features are More-Good, More-Bad, Less-Good, Less-Bad [11]. Two other features are lexicon related derived from Leaman’s [5] lexicon, the two features in question are the Boolean feature of presence or absence of ADR mention, and the numeric feature of the number of times an ADR is mentioned, they also collect a topic based feature using Mallet tool [11] [49], the features are the topic mentioned and the sum of the relevance score [11]. Other features include: length of the text, the presence of certain tags from Stanford parser. To implement Naive Bayes (NB) and Maximum Entropy (ME) Weka tools are used, for SVM they use LibSVM implementation [11]. The SVM uses RBF kernel. Several combinations of the dataset were used to test the program, giving varying results, but the after combining all the datasets and testing them on the test data, it outperformed all the others, the results are shown in this table [11].

Also not all features have shown improvement as indicated by this table [11]. Where removing a feature tends to cause a drop in performance, except Synonyms (Syn-set) and topic model, have limited to no effect on the datasets (Except ADE corpus) [11].



[12] Tried to not only identify ADRs but drug users, implementing two SVM classifiers, one to first identify users of five cancer drugs (still on trial) on Twitter , and the other to identify which ADR did they suffer [12]. The dataset acquired was is a 2 billion tweet collection collected using the Twitter API collected by [50] [51] for a different study [12]. For this study, the Tweets were reduced to four fields, 1) ID, 2) User ID, 3) Timestamp, and 4) text. The features extracted from the Tweets are textual and semantic [12], textual features include Bag-of-Words (BoWs), number of hasht-tags in the document, number of reply tags, number of negating words, the number of URLs, the number of pronouns, and the number of occurrences of the drug names or synonyms [12]. The semantic features are those derived from the UMLS [31] concepts extracted using MetaMap [43]. Trying to limit the data by choosing specific keywords more specifically the number of semantic type and group. After some trials, the Apache Lucene information retrieval library [52] parallelized with Amazon Cloud E2C was used. This lead to having the dataset limited to 239 potential users of the drug, 72 of them were confirmed. The evaluation of the classifiers was vaguely defined, but what is clear is that the first classifier had a prediction accuracy of 0.74 and an Area-Under the-Curve (AUC) of 0.82, and the second got an accuracy of 0.74 and AUC of 0.74 [12].

In [13], the works of the previous study was criticized over the technical aspect of using 72 sized dataset for both training and testing and not using the classifiers to classify more raw Tweets [13], it also criticises the lack of mentioning for the overall results, as well as using investigational drugs (still on trials) as it is not recognized whether the tweeter is using placebos or the real drugs [13]. With that in mind, the drugs chosen (Duloxetine, Gabapentin, Baclofen, Glatiramer, Pregabalin) have been chosen for being in the market for a number of years. This time the Tweets were mined manually of the period of 80 days using Twitter API (Which does not allow searching for posts older than 2 weeks) for a total of 6829 tweets, removing matching brand names of the drugs to peoples’ names and using only English tweets [13]. The features used were personal pronouns and sentiments derived from the NLTK [47], this is because the study required having “personal experience” tweets as opposed to “non-personal” tweets, according to the study; personal experience is expressed more often with pronouns. Using this principal, three classifiers were made, Naïve Bayes, SVM and Maximum Entropy (ME) [13]. ME was the most superior at precision= 0.866, Recall=0.842 and F-Measure=0.848, SVM’s result was precision= 0.856, Recall=0.810 and F-Measure=0.820, and Naïve Bayes was precision= 0.858, Recall=0.827 and F-Measure=0.835. Using 600 tweets for training and 285 for testing [13].

**Hybrid and Unique**

Of mixed solutions, a mixed unsupervised and supervised learning [14] system (called ADRMine) had been developed. Data was provided by DailyStrength [25] and twitter using 81 drugs. The system used the same data provided in [24] which was used by [11]. Expert annotators annotated the posts and matched it with IAA’s Cohen’s Kappa [46], with the gold standard includes only reviews with complete IAA. The result is 4720 reviews from DailyStrength (+1559 test) and 1340 tweets (+444 test). An additional 313833 DS reviews and 397729 drug related tweets were gathered in a total 711562 postings to form an unlabelled set which would be used in unsupervised learning. A lexicon was generated using the previously mentioned tools UMLS [31] and SIDER [14] [32]. Concepts are extracted using Conditional Random Fields (CRF), a CRF classifier is used to extract ADR concepts from user sentences using CRFsuite implementation [53], turning them into individual tokens, beneficial effects were also identified as it was noted that including them improves performance of ADR extraction [14]. The CRF features extracted include context features: the 3 tokens before 3 tokens after and current token (Spelling correction was done with Apache Lucene [52]). ADR lexicon: a binary feature that shows whether or not current token is included in the lexicon [14]. POS: generated with Stanford parser. Negation: Features that indicated that the token is negated using syntactic dependency rule [14]. The other feature extracted is the learning word embedding, the embedding is a meaningful real-valued vector of configurable dimension (between 50 and 500), these vectors were generated using Word2vec tool [54], which learns the embedding based on the word’s contexts in different sentences, then a K-Mean clustering operation is performed to cluster the words into n (=150) different clusters (Each cluster has some common words, like one only including ADR or only including drug names or dates) [14]. Seven features are defined based on the generated clusters, which are the cluster numbers of the current token, the three preceding and the three following tokens. ADRMine’s CRF classifier was compared to 4 extraction techniques (SVM, Lexicon-based, and two simple baselines based on MetaMap). The system proved superior to the other techniques, reaching an F-Score= 0.821 for a recall=0.784 and precision=0.860 for DS dataset and F-Score =0.721 for Recall=0.682 and Precision=0.765 for twitter. It was concluded that the lexicon, POS and negation features did not add a significant contribution with a huge dataset as it did with smaller dataset [14].

There was only one example of hidden Markov Model [15], the study describes three type of sources for ADR information, biomedical sources (books, journals, magazines, drug package labels) which may not be up to date, most accurate [15]. Clinical sources (patients’ data) which are not free and are limited by ethical, legal and social constraints, may also be inaccurate. Online Forums are the most inaccurate, but they are the most numerous, up to date and totally free, refers to many other works found in the survey [15]. The data was extracted from Medication.com [28] and steadyHealth.com [55] using JSoup crawler, collecting 8065 posts from Medication and 11878 from SteadyHealth [15]. To extract relationships between entities, the information extraction module is made, consisting of Named Entity Recognition (NER) and Relationship Extraction (RE) sub-modules [15]. NER helps to identify entities of interest in a given text, such as names of drugs, side effects and keywords or phrases relating them together using lexicon based method [15]. The dictionary of drug names was crawled from the drug lists on drugs.com [56], side effects from SIDER [32]. RE than identifies the relationships between named entities using Hidden Markov Model, which learns the association between the drug name and side-effect in a given text [15]. The HMM is defined using the following parameters: N number of states, M number of observation symbols, A N by N transition probability matrix, B N by M observation probability matrix. Pi N by 1 initial state probability vector. Around 2000 annotated training data messages is used to train the classifier using the Baum-Welch training algorithm [15]. The learnt model is used with Viterbi decoding algorithm to predict the hidden states for the observed sequence data in the testing set, where if the three states (drug, side-effect, connecting keyword) then the text is flagged as a positive drug/side-effect relationship. Using 10-fold-cross validation, the model got an F-score = 0.76, HMM however was unable to distinct between ADRs and symptoms of the drug, reducing the dictionaries themselves causes problems, especially for the drug dictionary [15].

Another unique approach was the Partially Supervised Learning technique (PSL) [16], which tries to solve the lack of large labelled data, by only giving the classifier a small number of labelled data and dynamically augmented it throughout the learning process, this would eliminate the need for a large expertly annotated dataset [16]. The dataset in question I extracted from Yahoo! Forums ProzacAwareness (Prozac drug) and Selective Serotonin Reuptake Inhibitors (SSRIsex), having a total of 6400 posts (1600 ADR and 4800 Non-ADR posts) [16]. A consensus detection is used to identify which new example goes to which cluster of data (positive or negative) using Rocchio algorithm [16]. An SVM and Naïve Bayes (NB) classifiers were created with and without PSL for comparison. SVM and NB scored less than 68 F-score without PSL, while the SVM and NB with PSL reach 89.74 and 86.32 F-score respectively [16]. The proposed approach was compared using benchmark labelling heuristics (EAT and PNLH) outperforming both of them in terms of F-Score [16].

# Work Plan

Initial Gantt chart for the work plan.

# References

|  |  |
| --- | --- |
| [1] | R. Edwards and F. K. Aronson, "Adverse Drug Reactions: Definitions, Diagnosis and Management," *Lancet,* vol. 356, no. Number 9237, pp. 1255-1259, 7 October 2000. |
| [2] | World Health Organization, “The Importance of Pharmacovigilance-Safety Monitoring of Medicinal Products,” 2002. [Online]. Available: http://apps.who.int/medicinedocs/en/d/Js4893e/1.html. [Accessed 18 11 2018]. |
| [3] | A. Nikfrajam, K. Oconnor, A. Sarker, R. Ginn, S. Jayaraman, T. Upadhya and G. Gonzalez, “Utilizing social media data for pharmacovigilance: A review,” *Research Gate,* vol. 54, pp. 202-212, 2015. |
| [4] | US Food & Drug Administration (FDA), “Preventable Adverse Drug Reactions: A Focus on Drug Interactions,” US Food & Drug Administration (FDA), 6 3 2018. [Online]. Available: https://www.fda.gov/drugs/developmentapprovalprocess/developmentresources/druginteractionslabeling/ucm110632.htm. [Accessed 19 11 2018]. |
| [5] | R. Leaman, L. Wojtulewicz, R. Sullivan, A. Shariah, J. Yang and G. Gonzalez, “Towards Internet-Age Pharmacovigilance: Extracting Adverse Drug,” in *Proceedings of the 2010 workshop On biomedical Natural Language Processing*, Uppsala, Sweden , 2010. |
| [6] | A. Nikfarjam and G. H. Gonzalez, “Pattern Mining for Extraction of mentions of Adverse Drug Reactions From User Comments,” *AMIA Annual Symposium Proceedings Archive,* vol. 2011, pp. 1019-1026, 2011. |
| [7] | D. Pappel, “Up To 90 Percent Of Adverse Reactions Unreported, Pharma Companies Search Social Media To Make Drugs Safer,” Inquisitr, 14 3 2015. [Online]. Available: https://www.inquisitr.com/1914900/up-to-90-percent-of-adverse-reactions-unreported-pharma-companies-search-social-media-to-make-drugs-safer/. [Accessed 19 11 2018]. |
| [8] | R. Sloane, O. Osanlou, D. Lewis, D. Bollegala, S. Maskell and M. Pirmohamed, “Social media and pharmacovigilance: A review of the opportunities and challenges,” *British Journal of Clinicla Pharmacology,* vol. 80, no. 4, pp. 910-920, 2015. |
| [9] | C. C. Yang and H. Yang, “Exploiting Social Media With Tensor Decomposition For Pharmacovigilance,” in *2015 IEEE 15th International Conference on Data Mining Workshops*, Atlantic City New Jersey, 2015. |
| [10] | J. Hadzi-Puric and J. Grumsa, “Automatic Drug Adverse Reaction Discovery From Parenting websites using disproportionality methods,” in *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Istanbul, 2012. |
| [11] | A. Sarker and G. Graciela, “Portable Automatic Text Classification for Adverse Drug Reaction Detection Via Multi-Corpus Training,” *Journal of biomedical informatics,* vol. 53, pp. 196-207, 2014. |
| [12] | J. Bian, U. Topaloglu and F. Yu, “Towards Large-Scale Twitter Mining For Drug-Related Adverse Events,” in *Proceedings of the 2012 ACM International Workshop on Smart Health and Wellbeing*, Hawaii, 2012. |
| [13] | Y. Zheng and K. Jiang, “Mining Twitter Data for Potential Drug Effects,” in *Advanced Data Mining and Applications*, Hangzhou, Springer, 2013, pp. 434-443. |
| [14] | A. Nikfarjam, A. Sarker, K. OConnor, R. Ginn and G. Gonzalez, “Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features,” *J Am Med Inform Assoc,* vol. 22, no. 3, pp. 671-681, 2015. |
| [15] | H. Sampathkumar, B. Luo and X. W. Chen, “Mining Adverse Drug Reactions from online healthcare forums using Hidden Markov Model,” *BMC Medical informatics and Decision Making,* vol. 14, no. 91, 2014. |
| [16] | M. Yang, X. Wang and M. Kiang, “IDENTIFICATION OF CONSUMER ADVERSE DRUG REACTION MESSAGES ON SOCIAL MEDIA,” in *PACIS 2013*, 2013. |
| [17] | J. Brownlee, “Classification Accuracy is Not Enough: More Performance Measures You Can Use,” Machine Learning Mastery, 21 3 2014. [Online]. Available: https://machinelearningmastery.com/classification-accuracy-is-not-enough-more-performance-measures-you-can-use/. [Accessed 25 11 2018]. |
| [18] | Google, “Classification: Precision and Recall,” Google, 1 10 2018. [Online]. Available: https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall. [Accessed 25 11 2018]. |
| [19] | W. Koehrsen, “Beyond Accuracy: Precision and Recall,” Towards Data Science, 3 3 2018. [Online]. Available: https://towardsdatascience.com/beyond-accuracy-precision-and-recall-3da06bea9f6c. [Accessed 25 11 2018]. |
| [20] | US Food & Drug Administration (FDA), “Questions and Answers on FDA's Adverse Event Reporting System (FAERS),” US Food & Drug Administration (FDA), 4 6 2018. [Online]. Available: https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/. [Accessed 20 11 2018]. |
| [21] | US Food & Drug Administration (FDA), “MedWatch: The FDA Safety Information and Adverse Event Reporting Program,” US Food & Drug Administration (FDA), 14 11 2018. [Online]. Available: https://www.fda.gov/Safety/MedWatch/default.htm. [Accessed 20 11 2018]. |
| [22] | S. Fox, “The Social Life of Health Information,” 12 5 2011. [Online]. Available: http://www.pewinternet.org/2011/05/12/the-social-life-of-health-information-2011/. [Accessed 20 11 2018]. |
| [23] | S. Vilar, C. Friedman and G. Hripcsak, “Detection of drug–drug interactions through data mining studies using clinical sources, scientific literature and social media,” *Briefings in Bioinformatics,* vol. 19, no. 5, p. 863–877, 28 9 2018. |
| [24] | Arizona State University, “TWITTER ANNOTATED CORPUS,” Arizona State University, 2014. [Online]. Available: http://diego.asu.edu/downloads/twitter\_annotated\_corpus/. [Accessed 20 11 2018]. |
| [25] | DailyStrength, “DailyStrength: Getting Better Together,” DailyStrength, 2006. [Online]. Available: https://www.dailystrength.org/. [Accessed 20 11 2018]. |
| [26] | MedHelp, “MedHelp: Be your healthiest,” MedHelp, [Online]. Available: https://www.medhelp.org/. [Accessed 20 11 2018]. |
| [27] | PatientsLikeMe, “PatientsLikeMe: Living better starts here.,” PatientsLikeMe, 2005. [Online]. Available: https://www.patientslikeme.com/. [Accessed 20 11 2018]. |
| [28] | Medications.com, “The Premier Community to talk about Health,” [Online]. Available: http://www.medications.com/. [Accessed 2 12 2018]. |
| [29] | Ask A patient, “When you need to know if a medication really works, why not Ask a Patient?,” [Online]. Available: https://www.askapatient.com/. [Accessed 2 12 2018]. |
| [30] | E. K. Ikonomakis, T. V. and K. Sotiris, “Text Classification Using Machine Learning,” *WSEAS Transactions on Computers,* vol. 4, no. 8, pp. 966-974, 2005. |
| [31] | U.S. National Library of Medicine, “Metathesaurus,” U.S. National Library of Medicine, 12 4 2016. [Online]. Available: https://www.nlm.nih.gov/research/umls/knowledge\_sources/metathesaurus/. [Accessed 28 11 2018]. |
| [32] | SiderEffects, “SIDER 4.1 : Side Effect Resource,” SiderEffects, 21 10 2015. [Online]. Available: http://sideeffects.embl.de/. [Accessed 28 11 2018]. |
| [33] | Government of Canadian, “MedEffect Canada,” Government of Canadian, 26 11 2018. [Online]. Available: https://www.canada.ca/en/health-canada/services/drugs-health-products/medeffect-canada.html. [Accessed 28 11 2018]. |
| [34] | Tartarus, “Snowball,” Tartarus, 2 2002. [Online]. Available: http://snowball.tartarus.org/. [Accessed 28 11 2018]. |
| [35] | Rosettacode, “Jaro distance,” Rosettacode, 29 7 2018. [Online]. Available: https://rosettacode.org/wiki/Jaro\_distance. [Accessed 28 11 2018]. |
| [36] | B. Liu, W. Hsu and Y. Ma, “Integrating Classification and Association Rule Mining,” in *The Fourth International Conference on Knowledge Discovery and Data Mining.*, New York, 1998. |
| [37] | The Stanford NLP Group, “Software > Stanford Log-linear Part-Of-Speech Tagger,” The Stanford NLP Group, 16 10 2018. [Online]. Available: https://nlp.stanford.edu/software/tagger.shtml. [Accessed 28 11 2018]. |
| [38] | Princeton University, “WordNet A Lexical Database for English,” Princeton University, 2018. [Online]. Available: https://wordnet.princeton.edu/. [Accessed 2018]. |
| [39] | Christian Borgelt's Web Pages, “Apriori - Association Rule Induction / Frequent Item Set Mining,” Christian Borgelt's Web Pages, 18 10 2018. [Online]. Available: http://www.borgelt.net/apriori.html. [Accessed 28 11 2018]. |
| [40] | A. C. Egberts, R. H. Meyboom and E. P. van Puijenbroek, “Use of Measures of Disproportionality in Pharmacovigilance,” *Spriner Link,* vol. 25, no. 6, pp. 453-458, 2002. |
| [41] | Serbian Government, “Medecines and Medical Devices Agency of Serbia,” Serbian Government, [Online]. Available: https://www.alims.gov.rs/eng/. [Accessed 2 12 2018]. |
| [42] | DrugBank, “The DrugBank database is a unique bioinformatics and cheminformatics resource that combines detailed drug data with comprehensive drug target information.,” DrugBank, [Online]. Available: https://www.drugbank.ca/. [Accessed 2 12 2018]. |
| [43] | USA government, “MetaMap - A Tool For Recognizing UMLS Concepts in Text,” USA government, 23 1 2018. [Online]. Available: https://metamap.nlm.nih.gov/. [Accessed 29 11 2018]. |
| [44] | B. W. Chee, R. Berlin and B. Schatz, “Predicting Adverse Drug Events from Personal Health Messages,” in *AMIA Annual Symposium Proceedings Archive*, Washington DC, 2011. |
| [45] | MedDRA, “Welcome to MedDRA,” MedDRA, [Online]. Available: https://www.meddra.org/. [Accessed 1 12 2018]. |
| [46] | Brandeis University, “Inter Annotator Agreement,” Word Press, 28 2 2017. [Online]. Available: https://corpuslinguisticmethods.wordpress.com/2014/01/15/what-is-inter-annotator-agreement/. [Accessed 29 11 2018]. |
| [47] | Sphinx, “Natural Language Toolkit,” Sphinx, 17 11 2018. [Online]. Available: http://www.nltk.org/. [Accessed 29 11 2018]. |
| [48] | Carnegie Mellon, “Tweet NLP,” Carnegie Mellon, [Online]. Available: http://www.cs.cmu.edu/~ark/TweetNLP/. [Accessed 29 11 2018]. |
| [49] | University of Massachusetts Amherst, “MAchine Learning for LanguagE Toolkit,” University of Massachusetts Amherst, 2002. [Online]. Available: http://mallet.cs.umass.edu/index.php. [Accessed 29 11 2018]. |
| [50] | B. O'Connor, R. Balasubramanyan, B. R. Routledge and N. A. Smith, “From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series,” in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 2010. |
| [51] | M. J. Paul and M. Dredze, “You are what you Tweet: Analysing Twitter fo public Health,” in *file:///C:/Users/Ahmed/Downloads/2880-14200-1-PB.pdf*, Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011. |
| [52] | Apache, “Apache Lucene Core,” Apache, 2011. [Online]. Available: https://lucene.apache.org/core/. [Accessed 1 12 2018]. |
| [53] | N. Okazaki, “CRFsuite A fast implementation of Conditional Random Fields (CRFs),” Naoaki Okazaki's website, 25 5 2016. [Online]. Available: http://www.chokkan.org/software/crfsuite/. [Accessed 2 12 2018]. |
| [54] | Google, “word2vec,” Apache, 30 7 2013. [Online]. Available: https://code.google.com/archive/p/word2vec/. [Accessed 2 12 2018]. |
| [55] | SteadyHealth, SteadyHealth, [Online]. Available: https://www.steadyhealth.com/. [Accessed 2 12 2018]. |
| [56] | Drugs.com, “Worried about drug interactions? Use the Interactions Checker,” Drugs.com, [Online]. Available: https://www.drugs.com/. [Accessed 2 12 2018]. |
| [57] | Healthcare Triage, “How Does the FDA Approve a Drug?,” The Incidental Economist, 20 Jul 2015. [Online]. Available: https://www.youtube.com/watch?v=WUsO6PH\_O54. [Accessed 2 3 2019]. |