Sentiment Analysis of Drug Reviews using Machine Learning Algorithms

Harika Kudarvalli (0887033)
Department of Computer Science
Lakehead University
Thunder Bay, Canada
hkudarva@lakeheadu.ca

Dr. Thiago E. A. Oliveira
Department of Computer Science
Lakehead University
Thunder Bay, Canada
thiago.oliveira@lakeheadu.ca

***Abstract*—With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges are rising exponentially. The abundance of data has made it easy for data scientists to analyse and draw conclusions about public's opinions and behaviour. Sentiment analysis is contextual mining of text which identifies and extracts subjective information in source material, and helps a business to understand the social sentiment of their brand, product or service while monitoring online conversations. We have briefly discussed about sentiment analysis and its related work in the domain from all around the world. The dataset being used is called "Drug Review Dataset" which contains patient's reviews on various drugs. We have applied multiple machine learning algorithms to build our classifier namely, Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) and Random Forest. Their results in terms of accuracy and other performance metrics have been discussed. Lastly, we have evaluated performance of all classifiers and have concluded that LSTM achieves best accuracy among other classifiers.**

***Index Terms*—Sentiment Analysis, Drug Reviews, Machine Learning. Health Recommender System.**

I. Introduction

Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. A sentiment analysis [11] system for text analysis combines natural language processing (NLP) and machine learning techniques to assign weighted sentiment scores to the entities, topics, themes and categories within a sentence or phrase. Sentiment analysis helps data analysts within large enterprises gauge public opinion, conduct nuanced market research, monitor brand and product reputation, and understand customer experiences.The primary role of machine learning in sentiment analysis is to improve and automate the low-level text analytics functions that sentiment analysis relies on, including Part of Speech tagging. For example, data scientists can train a machine learning model to identify nouns by feeding it a large volume of text documents containing pre-tagged examples. Using supervised and unsupervised machine learning techniques, such as neural networks and deep learning, the model will learn what nouns look like.

Pharmaceutical product safety currently depends on clinical trials and specific test protocols. Such studies are typically done under standardized conditions in a limited number of test subjects within a limited time span. Therefore, post-marketing drug surveillance, i.e. pharmacovigilance, plays a major role concerning drug safety once a drug has been released. Sentiment analysis of patient data in general and on drug experience in particular is a challenging research problem that is currently receiving considerable attention. One of the main issues, however, is the lack of annotated data, which is crucial for accurate sentiment classification. Especially, labeled data dealing with distinct aspects is rare. Moreover, the availability of labeled data is highly domain dependent. Therefore, in recent years, research efforts have shifted toward developing cross-domain techniques to solve this issue.
The goal of this research was to understand how our dataset has been distributed and in what ways we can apply sentiment analysis on it. The performance of multiple machine learning algorithms has been compared in accordance to their accuracies.

This paper is outlined as follows. **Section II** narrates the related work. **Section III** discusses the methodology used to develop the models, dataset used and various machine learning algorithms. **Sections IV** discuss the results. **Section VII** concludes the work.

II. Related Work

A. Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and Cross-Data Learning

Online review sites and opinion forums contain a wealth of information regarding user preferences and experiences over multiple product domains. Authors [6] examine online user reviews within the pharmaceutical field. Online user reviews in this domain contain information related to multiple aspects such as effectiveness of drugs and side effects, which make automatic analysis very interesting but also challenging. However, analyzing sentiments concerning the various aspects of drug reviews can provide valuable insights, help with decision making and improve monitoring public health by revealing collective experience. Through this paper, we can learn the transfer learning approaches that can be used to exploit similarities across domains and is a promising approach for cross-domain sentiment analysis.

B. Sentiment lexicons for health-related opinion mining

Opinion mining consists in extracting from a text opinions expressed by its author and their polarity. Lexical resources, such as polarized lexicons, are needed for this task. Opinion mining in the medical domain has not been well explored, partly because little credence is given to patients and their opinions (although more and more of them are using social media). The authors of [4] present the creation of our lexical resources and their adaptation to the medical domain and some words have a different polarity in the general domain and in the medical one. With the help of this paper, we can evaluate the lexicons and show with a simple algorithm that using our general lexicon gives better results than other well-known ones on our corpus and that adding the domain lexicon improves them as well.

C. Approaches to Cross-Domain Sentiment Analysis: A Systematic Literature Review

A sentiment analysis has received a lot of attention from researchers working in the fields of natural language processing and text mining. Many research studies have attempted to tackle this issue and to improve cross-domain sentiment classification. Results [1] of a comprehensive systematic literature review of the methods and techniques have been employed in a cross-domain sentiment analysis. Hence, one of the aims of this review is to create a resource in the form of an overview of the techniques, methods, and approaches that have been used to attempt to solve the problem of cross-domain sentiment analysis in order to assist researchers in developing new and more accurate techniques in the future.

D. Patient opinion mining to analyze drugs satisfaction using supervised learning

Opinion mining is a very challenging problem, since user generated content is described in various complex ways using natural language. In opinion mining, most of the researchers have worked on general domains such as electronic products, movies, and restaurants reviews but not much on health and medical domains. Few studies investigating the impact of social media on patients have shown that for some health problems, online community support results in a positive effect. The opinion mining method employed in this work focuses on predicting the drug satisfaction level among the other patients who already experienced the effect of a drug. The results in [5] demonstrate that neural network-based opinion mining approach outperforms the support vector machine method in terms of precision, recall and f-score. It is also shown that the performance of radial basis function neural network method is superior than probabilistic neural network method in terms of the performance measures used.

E. Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts

Based on the intuition that patients post about Adverse Drug Reactions (ADRs) expressing negative sentiments, we investigate the effect of sentiment analysis features in locating ADR mentions. The authors focus the feature space of a state-of-the-art ADR identification method with sentiment analysis features. Using a corpus of posts from the DailyStrength forum and tweets annotated for ADR and indication mentions, we can evaluate the extent to which sentiment analysis features help in locating ADR mentions and distinguishing them from indication mentions. Evaluation results show that sentiment analysis features marginally improve ADR identification in tweets and health related forum posts. This study [10] shows that adding sentiment analysis features can marginally improve the performance of even a state-of-the-art ADR identification method. This improvement can be of use to pharmacovigilance practice, due to the rapidly increasing popularity of social media and health forums.

	uniqueID	drugName	condition	review	rating	date	usefulCount
0	206461	Valsartan	Left Ventricular Dysfunction	"It has no side effect, I take it in combinati...	9	2012-05-20	27
1	95260	Guanfacine	ADHD	"My son is halfway through his fourth week of ...	8	2010-04-27	192
2	92703	Lybrel	Birth Control	"I used to take another oral contraceptive, wh...	5	2009-12-14	17
3	138000	Ortho Evra	Birth Control	"This is my first time using any form of birth...	8	2015-11-03	10
4	35696	Buprenorphine / naloxone	Opiate Dependence	"Suboxone has completely turned my life around...	9	2016-11-27	37

Fig. 1. Data Description

III. Methodology

The entire experiment has been defined by following steps:

1) Data Retrieval (Dataset)
2) Data Pre-processing
3) Data Cleaning
4) Data Visualization
5) Feature Extraction
6) Machine Learning Algorithms
7) Training Testing Model
8) Evaluation Metrics
9) Comparison

A. Data Retrieval (Dataset)

The dataset being used has been extracted from user reviews and ratings of two webpages – Drugs.com Drugslib.com. Drugs.com allows users to review specific drugs with condition and 10-star rating. Whereas the later has limited ratings but in a very structured format. Reviews have been grouped into three aspects – benefits, side effects and overall comment. The authors [6] of this dataset have gathered data using automatic web crawlers. Beautiful Soup library in python was used for scraping data from html pages. It comprises of 215063 reviews from Drugs.com and 3551 reviews from Druglib.com. Fig. 8 depicts the first five entries of our training data. Later the dataset has been divided into training and testing data as 75% and 25% respectively. Drugs.com data have been selected from a defined list and thus standardized. Drugslib.com reviews have not been normalized but they consist of manifold variations. The average number of reviews per drug from Drugs.com is very high that in Drugslib.com

```
Pre - Processed Text:
 38961     "I've had paraguard since 2012 and I've not ha...
87028     "sooo, I started these pill back in January an...
20110     "This medicine has curbed my cravings for vodk...
77874     "I was super worried about trying Accutane. I ...
118902    "I had my first colonoscopy last week.  I was ...
Name: review, dtype: object
Processed Text:
 38961     paraguard since any pregnancy check string reg...
87028     started pill back january gained pound here la...
20110           medicine curbed craving sober day little no
77874     super worried trying heard strong that made sk...
118902    my first colonoscopy last prepared nearly bad ...
Name: review, dtype: object
```

Fig. 2. Pre - Processing

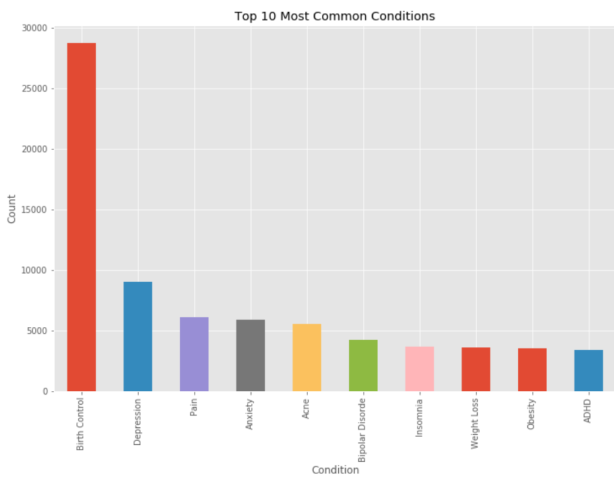

Fig. 3. Data Visualization

B. Data Pre-Processing

Raw data obtained from users have to be pre-processed before feeding it to a model. This increases accuracy and less performance time. There are many pre-processing stages [7] in text classification. In this experiment, we have converted all uppercase letters into lowercase letters. Stop words are removed too. Removal of unnecessary data reduces dataset but the intent is not impacted.

C. Data Cleaning

Once data is pre-processed, it has to go through a function to replace old data with cleaned data. Without performing this step the model might take uncleaned data as input.

D. Data Visualization

Once the data is cleaned and ready to be fed to the model, we can understand the data more closely with visualizing its factors in graphs or plots. Graphs give a clear picture about the dataset being dealt. Fig. 3 lists the top 10 most conditions present in our dataset. The cleaned data has already been split into training and testing sets.

E. Feature Extraction

Term-frequency-inverse document frequency (TF-IDF) is another way to judge the topic of an article by the words it contains. With TF-IDF [12], words are given weight – TF-IDF measures relevance, not frequency. That is, wordcounts are replaced with TF-IDF scores across the whole dataset.

F. Machine Learning Algorithms

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning [13] focuses on the development of computer programs that can access data and use it learn for themselves

1) Random Forest - Random forest [8] consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction. A large number of relatively uncorrelated models (trees) operating as a committee will outperform any of the individual constituent models.
2) Convolutional Neural Network - A great way to use deep learning to classify images is to build a Convolutional Neural Network (CNN). The Keras library in Python makes it pretty simple to build a CNN.CNN is a class of deep, feed-forward artificial neural networks (where connections between nodes do not form a cycle) & use a variation of multi layer perceptrons [2] designed to require minimal pre-processing. These are inspired by animal visual cortex.
3) Long Short-Term Memory - LSTMs were introduced to solve the vanishing gradient problem. They help preserve the error that can be back-propagated through time and layers. By maintaining a more constant error, they allow recurrent nets to continue to learn over many time steps (over 1000), thereby opening a channel to link causes and effects remotely. LSTMs [9] contain information outside the normal flow of the recurrent network in a gated cell. Information can be stored in, written to, or read from a cell, much like data in a computer's memory.

G. Training & Testing Model

Once the network of each algorithm is built, the data has to undergo training where we perform supervised learning. Algorithm uses train data to learn features from dataset. Later, testing is performed to evaluate the performance of our built model using performance metrics like accuracy, precision f1 score and recall. Also called as confusion matrix.

IV. Results

We have performed sentiment analysis on ”Drug Reviews” dataset and found some interesting insights. Almost 75% of entire dataset as shown in Fig. 3 seems to have positive reviews. The reviews have been divided into positive and negative response using features of reviews and 10 - star rating.

Our experiment started with training our dataset using Random Forest which is one of the most effective machine

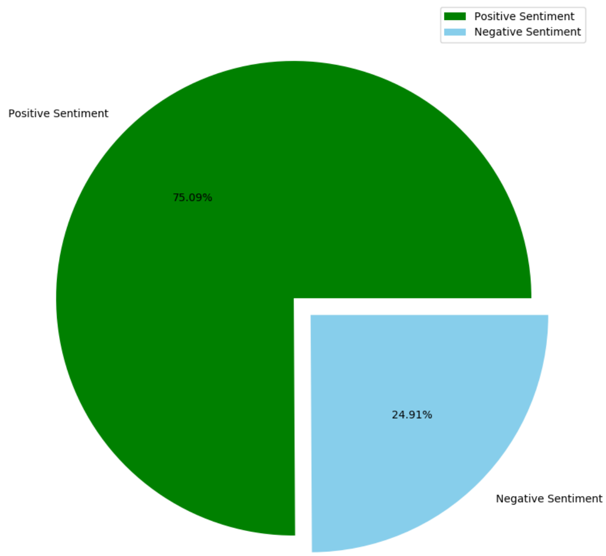

Fig. 4. Sentiment Distribution

	precision	recall	f1-score	support
0	0.33	0.71	0.45	9828
1	0.94	0.77	0.84	60150
accuracy			0.76	69978
macro avg	0.64	0.74	0.65	69978
weighted avg	0.86	0.76	0.79	69978

0.7574523421646804

Fig. 5. Classification Report for Random Forest

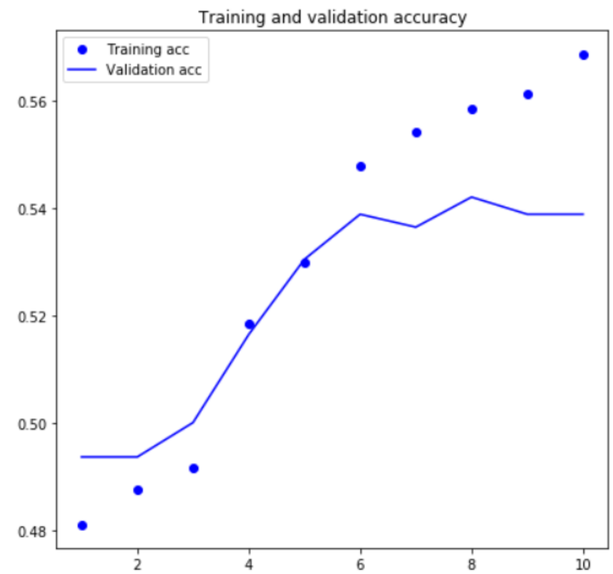

Fig. 6. Accuracy Curve for CNN

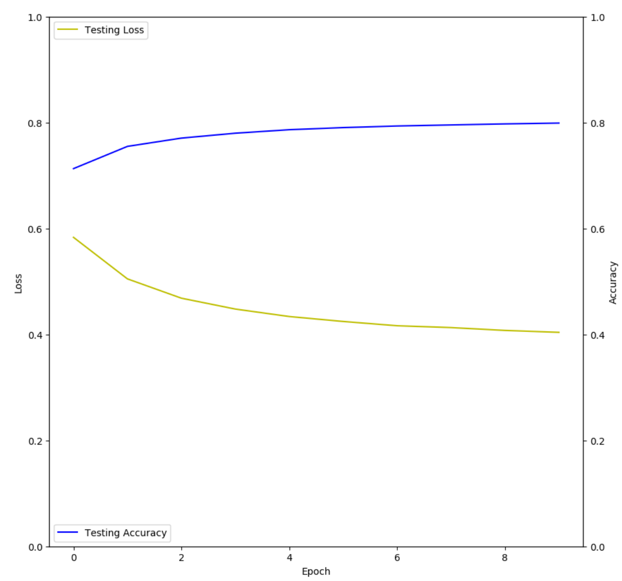

Fig. 7. Accuracy Curve for LSTM

learning algorithm. After training, testing and performing cross validation [14] using multiple parameters, we achieved an accuracy of 75.7%.

Further, a comparison with simple deep learning algorithm was observed. Convolutional neural network seems to perform with an accuracy of 58.04%. This is comparitvely less than Random Forest.

To increase the efficiency of our model, we trained our model using Long - Short Term Memory. It started with an accuracy of 80% and after training for 10 epochs, it was able to achieve an accuracy of - percentage. LSTM seems to do the best job out of all classifiers.

There is a slight difference in LSTM's and Random Forest's accuracy. The later works really well with tabulated or structured data [3]. Our dataset was tabulated with multiple columns filled with text. If our dataset consisted of unstructured data like images, audio or signals, LSTM would have performed way more efficiently than Random Forest.

V. Conclusion

With this research work, we have observed that machine learning algorithms play an important role in performing sentiment analysis. After performing pre-processing steps to our dataset, we were able to extract effective features. Multiple machine learning algorithms were performed on our dataset to understand their performances. From this research it is clear that no specific algorithm gives the best and accurate results. in conclusion, Deep learning algorithms perform better than machine learning algorithms. LSTM depicted best results among Random Forest and CNN. 10-star rating and the features from reviews were considered to evaluate the overall sentiment of our data. However with further contributions and tributes in the field of sentiment analysis, we will be able to achieve effective results with any type of dataset.

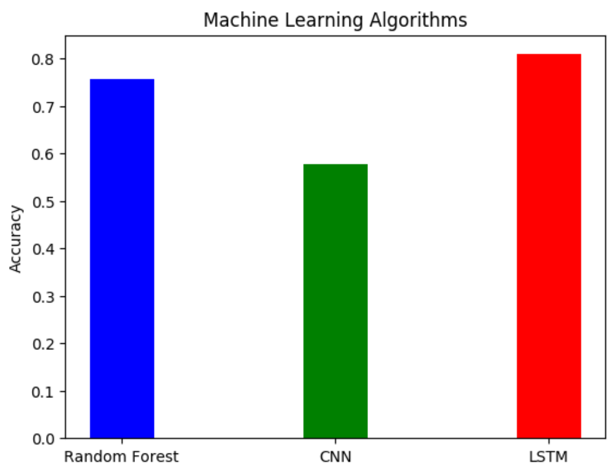

Fig. 8. Comparison among algorithms

VI. Future Work

We would like to extend our research in comparing advanced deep learning algorithms with unstructured data. A website can be designed as a repository for any drug information in the world. It can consists multiple reviews from other websites and some interactive graphical representations about data.

References

[1] Tareq Al-Moslmi, Nazlia Omar, Salwani Abdullah, and Mohammed Albared. Approaches to Cross-Domain Sentiment Analysis: A Systematic Literature Review. *IEEE Access*, 5:16173–16192, 2017.

[2] Ronan Collobert and Jason Weston. A unified architecture for natural language processing. pages 160–167, 2008.

[3] Robin Genuer, Jean-michel Poggi, Christine Tuleau-malot, Robin Genuer, Jean-michel Poggi, Christine Tuleau-malot, Nathalie Villavialaneix Random Forests, and Robin Genuer. Random Forests for Big Data. pages 28–46, 2017.

[4] Lorraine Goeuriot, Jin Cheon Na, Wai Yan Min Kyaing, Christopher Khoo, Yun Ke Chang, Yin Leng Theng, and Jung Jae Kim. Sentiment lexicons for health-related opinion mining. In *IHI'12 - Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, pages 219–225, 2012.

[5] Vinodhini Gopalakrishnan and Chandrasekaran Ramaswamy. Patient opinion mining to analyze drugs satisfaction using supervised learning. *Journal of Applied Research and Technology*, 15(4):311–319, 2017.

[6] Felix Gräßer, Hagen Malberg, Surya Kallumadi, and Sebastian Zaunseder. Aspect-Based sentiment analysis of drug reviews applying cross-Domain and cross-Data learning. *ACM International Conference Proceeding Series*, 2018-April:121–125, 2018.

[7] Vairaprakash Gurusamy and Subbu Kannan. Preprocessing techniques for text mining. 10 2014.

[8] Tin Kam Ho. Random decision forests. *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*, 1:278–282, 1995.

[9] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9:1735–80, 12 1997.

[10] Ioannis Korkontzelos, Azadeh Nikfarjam, Matthew Shardlow, Abeed Sarker, Sophia Ananiadou, and Graciela H. Gonzalez. Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. *Journal of Biomedical Informatics*, 62:148–158, 2016.

[11] Bing Liu. Sentiment analysis and subjectivity. *Handbook of Natural Language Processing, Second Edition*, pages 627–666, 2010.

[12] Nina Tahmasebi and Thomas Risse. On the uses of word sense change for research in the digital humanities. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10450 LNCS:246–257, 2017.

[13] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, and A. Mueller. Scikit-learn. *GetMobile: Mobile Computing and Communications*, 19(1):29–33, 2015.

[14] XueFei Zhou. Understanding the convolutional neural networks with gradient descent and backpropagation. *Journal of Physics: Conference Series*, 1004:012028, apr 2018.