VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
Department of Computer Engineering

Project Report on

MEDICAL REVIEW MINING USING SOCIAL MEDIA

In partial fulfillment of the Fourth Year, Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai Academic Year 2017-2018

Submitted by

Dipen Chawla, D17A 13
Disha Mohnani, D17A 41
Varsha Sawlani, D17A 64
Sujay Varma, D17A 79

Project Mentor

Mrs. Sujata Khedkar

(2018-19)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
Department of Computer Engineering

Certificate

This is to certify that ***Dipen Chawla, Disha Mohnani, Varsha Sawlani and Sujay Varma*** of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on "***Medical Review Mining Using Social Media***" as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor ***Prof. Sujata Khedkar*** in the year 2018-2019 .

This thesis/dissertation/project report entitled "***Medical Review Mining Using Social Media***" by ***Dipen Chawla, Disha Mohnani, Varsha Sawlani and Sujay Varma*** is approved for the degree of ***Bachelor of Engineering.***

Programme Outcomes	Grade
PO1,PO2,PO3,PO4,PO5,PO6,PO7, PO8, PO9, PO10, PO11, PO12 PSO1, PSO2	

Date:

Project Guide:

Project Report Approval
For
B. E (Computer Engineering)

This thesis/dissertation/project report entitled "***Medical Review Mining Using Social Media***" by ***Dipen Chawla, Disha Mohnani, Varsha Sawlani and Sujay Varma*** is approved for the degree of ***Bachelor of Engineering.***

Internal Examiner

External Examiner

Head of the Department

Principal

Date:
Place:

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

--	--
(Signature)	(Signature)
--	--
(Dipen Chawla, D17A 13)	(Disha Mohnani, D17A 41)
--	--
(Signature)	(Signature)
--	--
(Varsha Sawlani, D17A 64)	(Sujay Varma, D17A 79)

Date:

ACKNOWLEDGMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Mrs. Sujata Khedkar** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement at several times.

Computer Engineering Department

COURSE OUTCOMES FOR B.E PROJECT

Learners will be to,

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solution for the problem.
CO 4	Able to interpret the data and datasets to be utilized.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one's competency.
CO 11	Able to develop professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

ABSTRACT

Data mining combined with sentiment analysis is an ongoing field of research in the area of natural language processing using machine learning. It uses computational linguistics to systematically identify, extract and study of complex data. This report proposes an approach for developing a framework to analyze reviews from various medical blogs for drugs prescribed for neurological disorders. This would give an in-depth idea about the effectiveness of the drugs from the user's point of view. User reviews even provide the pros and cons of the drug and would also be helpful in generating an overall summary of a drug.

The development framework takes data from three medical websites viz webmd, drugs.com and everydayhealth.org.. On feeding this data to the preprocessing module would provide data which is ready to be annotated. Annotation module would annotate the data in order to categorize it into effective, ineffective, adverse reaction categories. Reviews not lying in any of these categories would be considered to belong to the 'others' category. Once annotated, the data is ready to be trained. We propose to use a semi-supervised approach for training the model. This approach involves a combination of machine learning and deep learning techniques for developing the model. The output would be generated in the form of reports and visualizations to provide the cons of the drug as well as to get the top sentences in each category. Report generation and visualization would be done by the visualization module.

The framework as a whole would be quite useful for both medical professionals as well as common users in general to obtain the effects of the drug as well as the comparison between them in order to suggest an alternative for drugs.

TABLE OF CONTENTS

CHAPTER INDEX

List of Figures

List of Tables

CHAPTER 1:

INTRODUCTION

In the introductory chapter, we have covered a brief introduction to our project, we have explained the motivation for our project, lacunas in the existing system, problem definition, relevance of the project and the methodology used.

1.1. Introduction to the project

Machine learning and Deep learning have a wide range of applications from product recommendations to produce a new drug and from producing medical diagnosis systems to fake news and music. They are widely being used in industries to give solutions to many problems. When it comes to effectiveness, ML and deep learning in pharmaceutical industries are glimmers of a potential future in which analysis and innovation are everyday scenarios. The use of machine learning in preliminary drug adverse reactions discovery has the potential for various uses, from initial screening of drug compounds to predicted market success rate based on biological factors.

1.2. Motivation for the project

Creating new drugs is very time consuming and costly process which requires a lot of efforts and analysis of related drugs. Adverse drug reactions are among the most common causes of death these days. Expensive clinical trials are not sufficient to uncover all of the adverse reactions a drug may cause, necessitating systems for monitoring the effects of medical drugs after they have been licensed for use. An enormous figure of researches and experiments are being done for new drug discoveries, the estimated US$2.6-billion price tag of developing a treatment. A lot of them apparently goes down the drain because money spent on nine out of ten treatments fail in phase I trial or regulatory phase of medical. Drug adverse reactions and effectiveness play an important role in discovering new or enhancing the previous drugs to give expected results for treatments. If the proponents of all these techniques are right, deep learning and machine learning can give quicker, cheaper and more effective drug discovery. Using these techniques, we will try to build a system which can help pharmaceutical industries to produce effective and cost friendly drugs for neurological disorders.

1.3 Problem Definition

Today's healthcare professionals use online platforms such as blogs, social media, and websites extensively to convey opinions on health matters and use of drugs. The use of such distributed information in an electronic medium, specifically the World Wide Web, with the ultimate aim to inform public health and public policy is known as ‘infodemiology’. On the other side of the healthcare spectrum, users commonly reach out to various social media platforms and blogs in order to convey their views about an administered drug. Extracting Adverse Drug Reactions (ADE) from such comments provides a new direction to the field of healthcare, which is already in the midst of a data explosion, thanks to the growth of social media as a tool for healthcare analysis. In the proposed project we will use deep learning approaches to analyze patient reviews on social media platforms about drugs they have been taking for their treatments and we will generate brief summaries and reports that can help the pharma industry for measuring efficiencies of drugs.

1.4 Existing Systems and its comparison table

Existing systems usually use the already available datasets. Most of the existing systems either use supervised machine learning approach or lexicon-based approach to develop the systems. Present systems classify medical drug reviews in only two categories viz. effective and ineffective. Also, the existing systems provide solutions for drugs used for commonly observed disorders.

Existing System	Our System
Binary classification of reviews.	Multiclass classification of reviews.
No adverseness comparison among drugs.	Adverseness comparison among drugs.
No comparison of effectiveness of each drug.	Compares effectiveness of each drug.
Do not give top rank sentences.	Gives top ranked sentences of each category.

1.5 Lacuna of the existing systems

Existing systems are incapable of providing an end-to-end solution for mining of drug reviews. These systems do not collect the reviews from the medical blogs or websites but use the readily available datasets. The supervised machine learning approach or lexicon-based approach used in existing systems reduces the scope of the project to deal with complex data. Classification of drug reviews into two categories does not give a clear idea about the cause of ineffectiveness or the severity of drug effects. Also, the existing systems do not provide a comparison between the effectiveness of multiple drugs for the same disorder. As a result, it sometimes becomes difficult to suggest alternative drugs for drugs having severe adverse effects.

1.6 Relevance of the Project

The developed project will be useful for both medical professionals and patients. The system generates an extractive summary of the drug under consideration by giving the top reviews in each category i.e effective, ineffective and adverse. It also generates visualizations in the form of adverseness comparison and the effectiveness of each drug. This would save the efforts of the patients required for checking multiple websites to get a brief about the drug administered to him/her. Besides, the user would also get an idea about the commonly experienced adverse effects for the drug. Based on the comparative report for the effects of drugs for the same disorder, medical practitioners can also suggest alternative drugs for the ones which are not quite effective. Also, as the system deals with generic names and not the brand names of the drugs, it treats drugs of different brands with the same composition as equal thereby ignoring the popularity of a particular brand of drug. The system developed thus tries to overcome all the drawbacks of existing systems proving to be more effective and useful.

1.7 Methodology employed for development

Agile Methodology:

We have followed continuous iterations of development and testing throughout the development lifecycle of the project. Both development and testing activities are concurrently practiced. Agile is a form of incremental and iterative approach to software design. Any error occurring during any phase of the project could be rectified immediately. Every module once developed was tested before moving on to the next phase. The entire team worked for every phase of the project.

Steps in agile methodology are:

1. Planning and Requirements Analysis- Requirement analysis is the most important and fundamental stage. This information is then used to plan the basic project approach.Planning for the quality assurance requirements and identification of the risks associated with the project is also done in the planning stage.
2. Design- A design approach clearly defines all the architectural modules of the product along with its communication and data flow representation with the external and third party modules. The internal design of all the modules of the proposed architecture should be clearly defined with the minutest of the details in designing.

3. Coding- In this stage the actual coding and implementation is performed. It involves writing the required code for implementation of all modules. Python programming language is majorly used for our system.
4. Testing and Integration- All the modules coded are tested and integrated. Testing is done at all levels eg unit testing, integration testing, etc. Once individual modules are ready they are then integrated to form the complete software.
5. Deployment- Once the software is tested and ready to be deployed it is released formally in the appropriate market.

CHAPTER 2:

LITERATURE SURVEY

In the literature survey, we have covered various research papers, books and papers referred, with their abstracts and the inference drawn.

2.1. Articles Referred

1. Using Data Science to help Women make Contraceptive Choices by Krittika Krishnan, Insight Data Science via Medium, October 19, 2018

Link:
https://blog.insightdatascience.com/using-data-science-to-help-women-make-contraceptive-choices-5e9cc4d8aac4

Abstract: The article showcases a webapp based on an intelligent NLP based classifier for contraceptive choice reviews.

Inference: The article brilliantly captures all the NLP techniques used for cleaning, preprocessing, feature extraction and sampling of reviews. Our project heavily relies on this work for aforementioned tasks as well as for the inspiration behind building the web application.

2. Using scikit-learn Pipelines and FeatureUnions by Zac Stewart

Link: http://zacstewart.com/2014/08/05/pipelines-of-featureunions-of-pipelines.html

Abstract: The article uses pipelines and sklearn feature unions for classification tasks.

Inference: The article was heavily used while combining and extracting features from text as well as combining them with VADER Sentiment scores.

3. Movie Review Sentiment Analysis EDA and models by Andrew Lukyanenko via Kaggle

Link: https://www.kaggle.com/artgor/movie-review-sentiment-analysis-eda-and-models

Abstract: This is a Python notebook tutorial for various classification model building as well as Exploratory Data Analysis.

Inference: The inspiration for building a OneVsRest classifier was obtained from this article as well as was helpful in the building of the LSTM models.

2.2. Research Papers

1. Kuhn, Michael, et al. "A side effect resource to capture phenotypic effects of drugs." ***Molecular systems biology*** **6.1 (2010): 343.**

Abstract: The paper introduces a public, computer-readable side effect resource (SIDER) that connects 888 drugs to 1450 side effect terms, as a method for studying Adverse Drug Reactions for academic and commercial research. The resource is freely available at http://sideeffects.embl.de.

Inference: The publicly available dataset is highly comprehensive in terms of number as well as quality of reviews and drug effects. The resource is potentially an important source of data for model testing and comparative analysis using lexicon approach.

2. Leaman, Robert, et al. "Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks." Proceedings of the 2010 workshop on biomedical natural language processing. Association for Computational Linguistics, 2010.

Abstract: Mining the relationships between drugs and adverse reactions of drugs from patient's opinion reviews mined from dailystrength website, and compares terms from the lexicon to user comments using a sliding window method.

Inference: The novel approach to pharmacovigilance by comparing comments to a lexicon dictionary is appreciated. The combined dataset is also immensely useful for the preparation of a corpus.

3. Nikfarjam, Azadeh, et al. "Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features." Journal of the American Medical Informatics Association 22.3 (2015): 671-681.

Abstract: This paper proposes a machine learning based sequence tagger for automatically extracting adverse drug reactions from user reviews. It uses a CRF based classifier which is trained using word embeddings performed on annotated reviews from a combination of comments from Twitter and dailystrength.com.

Inference: The model uses a highly efficient approach to obtain drug adverse reactions from comments using a combination of CRF classification technique and learning using word embeddings.

4. Yadav, Shweta, et al. "Medical Sentiment Analysis using Social Media: Towards building a Patient Assisted System." LREC. 2018.

Abstract: The paper introduces a novel approach for sentiment analysis to classify medical conditions and medication using data from social media and training a deep convolutional neural network to obtain sentiment word distribution for medical conditions and medication.

Inference: The five layered CNN model trained using dataset is highly accurate in terms of extracting sentiments from medical condition as well as medication reviews. The baseline models used for training are strong and provide necessary direction to actual training of model.

5. Huynh, Trung, et al. "Adverse drug reaction classification with deep neural networks." Coling, 2016.

Abstract: The paper establishes the classification models for drug reactions using two separate models, CRNN (a combination of Convolutional and Recurrent Neural Networks) and CNNA (Convolutional Neural Network with Attention). The classification is done on a Twitter dataset containing informal language using an ADE(Adverse Drug Reaction) dataset constructed from MEDLINE reports.

Inference: The neural networks constructed by this paper are extremely insightful and would go a long way in helping us to build our own models. The results are accurate as well for both of the deep learning models.

6. Yang, Christopher C., et al. "Social media mining for drug safety signal detection." Proceedings of the 2012 international workshop on Smart health and wellbeing. ACM, 2012.

Abstract: This paper uses association mining and proportional reporting ratios to mine relations between drugs and adverse reactions of drugs from user provided content from medhelp, an online medical community.

Inference : The novel idea of using association mining algorithm and lexicon for the classification of diseases and drug names is appreciated.

7. Lee, Kathy, et al. "Adverse drug event detection in tweets with semi-supervised convolutional neural networks." Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2017.

Abstract: This paper aims to provide a semi-supervised approach for classifying text and for detecting the adverse reactions of drugs in tweets collected.

Inference: The use of a semi-supervised CNN-based framework for classification of adverse drug events in tweets is considerably unique but is highly efficient as it produces better results than other state-of-the-art models trained on Twitter datasets.

8. Xia, Long, G. Alan Wang, and Weiguo Fan. "A Deep Learning Based Named Entity Recognition Approach for Adverse Drug Events Identification and Extraction in Health Social Media." International Conference on Smart Health. Springer, Cham, 2017.

Abstract: This paper introduces the use of natural language processing and deep learning for high performance and advanced extraction of adverse reactions of drugs from comments, from pubmed dataset as well as automatic crawlers for mining user comments.

Inference: The paper establishes the need for crawling newer comments rather than importing a dataset for review classification. Various adverse reactions are expected as results.

After researching the extensive work in the automated pharmacovigilance domain, we felt the need for building an accurate system that would be accessible and understandable to the end user for him/her to make an informed decision before consuming the aforementioned drugs.

CHAPTER 3:
REQUIREMENT GATHERING

In requirements for our system, we have explained the process of requirement gathering, functional requirements, non-functional requirements, constraints, hardware and software requirements, techniques, tools and algorithms used in the system.

3.1 Definition of requirement gathering

Requirements elicitation is the practice of researching and discovering the requirements of a system from users, customers, and other stakeholders.

3.2 Functional Requirements

3.2.1 Patient's point of view:

- Provide a measure of effectiveness, ineffectiveness and adverse nature of all the drugs.
- Should be able to cover the vast scope of terms in user reviews.
- Provide the major adverse reactions for the drug.

3.2.2 Medical professional's point of view:

- Should be able to compare different drugs for the same disorder in a visual as well as text format.
- Provide the pros and cons of the drug to be prescribed.
- Provide an overall summary of the prescribed drugs.
- Display the likelihood of having positive and negative effects of the drug.
- Predict whether a user review would be classified as adverse, effective or ineffective.

3.3 Non-Functional Requirements

- **Performance**

  The system gives efficient performance irrespective of the size and complexity of data.

- **Capacity**

  The system is able to handle huge amounts of data with ease.

- **Reliability**

  The system is able to perform specific functions without failure.

- **Scalability**

  The system is able to expand its processing capabilities upward and outward.

- **Availability**

  Users can depend on the system to be able to function during normal operating times.

- **Robust**

  The system is impervious to failures due to deployment.

- **Security and Privacy**

  The system is tolerant of bugs and can handle sensitive data with ease.

3.4. Constraints

- The model trained may not be 100% efficient.
- Some categories of classification may be neglected.
- Automated results may vary from original results.
- Less number of drugs are considered.
- Classification of a medical condition is not provided.
- Automated suggestions for other drugs are not given.

3.5.Various Hardware, Software, Technology and Tools available

Softwares : Anaconda Navigator, Orange, Weka, etc.

Technologies : nltk, spacy, numpy, matplotlib, plotly, bokeh, caffe, theano, pandas, keras, tensorflow, vader.

Tools : Jupyter Notebook, Spyder, Google Colab.

3.6 Selection of the Hardware, Software, Technology and Tools and their justification

1. Python 3 libraries: numpy, pandas, scrapy, nltk, spacy, matplotlib, textblob, etc. - Standard libraries for text and numerical analysis.
2. Scikit Learn - Model creation and result prediction.
3. Anaconda Navigator & Jupyter Notebook - Integrated environment for running python and notebook files.
4. TensorFlow and Keras for neural network - Libraries for neural network generation and simulation.
5. Google Colab - Shared environment with enhanced processing capabilities.
6. VADER for sentiment score calculation - Library for sentiment analysis of reviews.
7. Bokeh for dynamic visualisation - Visualisation library.

CHAPTER 4 :

PROPOSED DESIGN

In proposed design, we have covered the conceptual design, block diagram, modular diagram, data flow diagram, flow chart and Gantt chart.

4.1 Block diagram of the system

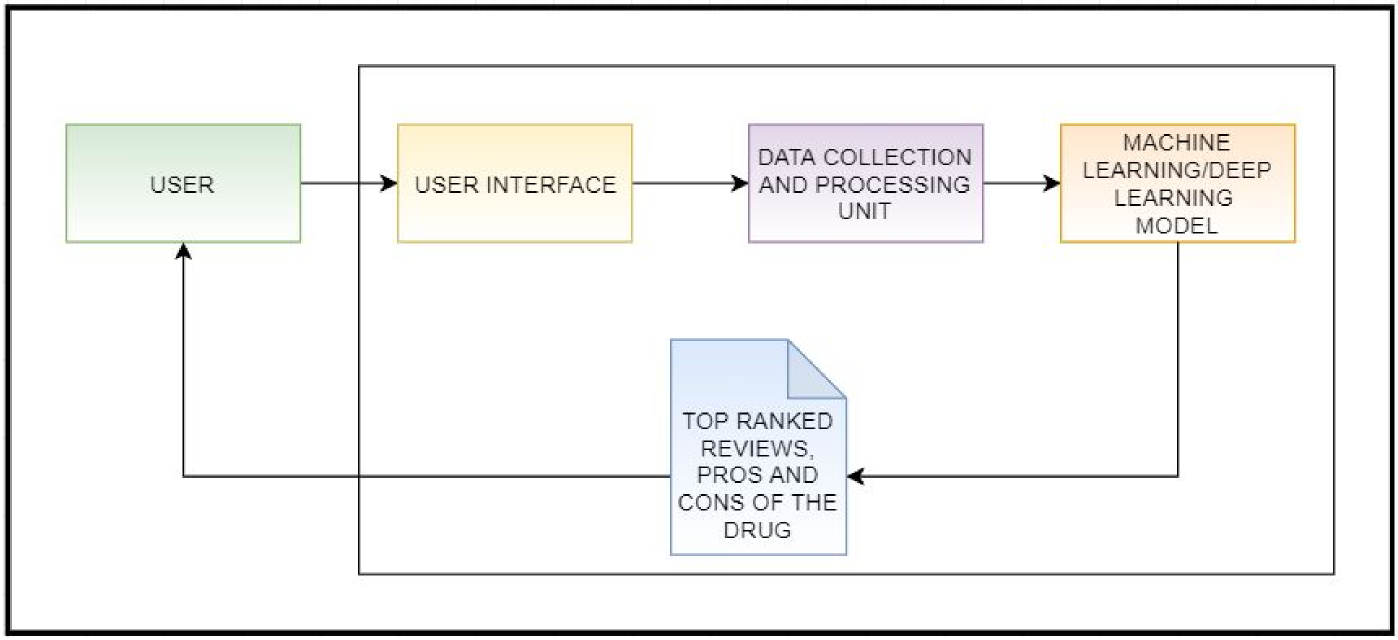


Fig 4.1 : Block diagram of the proposed system

As per the above block diagram, the reviews given by the users are collected and preprocessed. Preprocessing involves conversion of uppercase to lowercase, removal of stopwords and punctuation, stemming, lemmatization, normalization and tokenization.The clean data is then passed through our machine learning/deep learning model which gives classification of reviews into various categories. Based on this categorization, top ranked reviews, pros and cons of the drug are generated for the user.

4.2 Modular design of the system

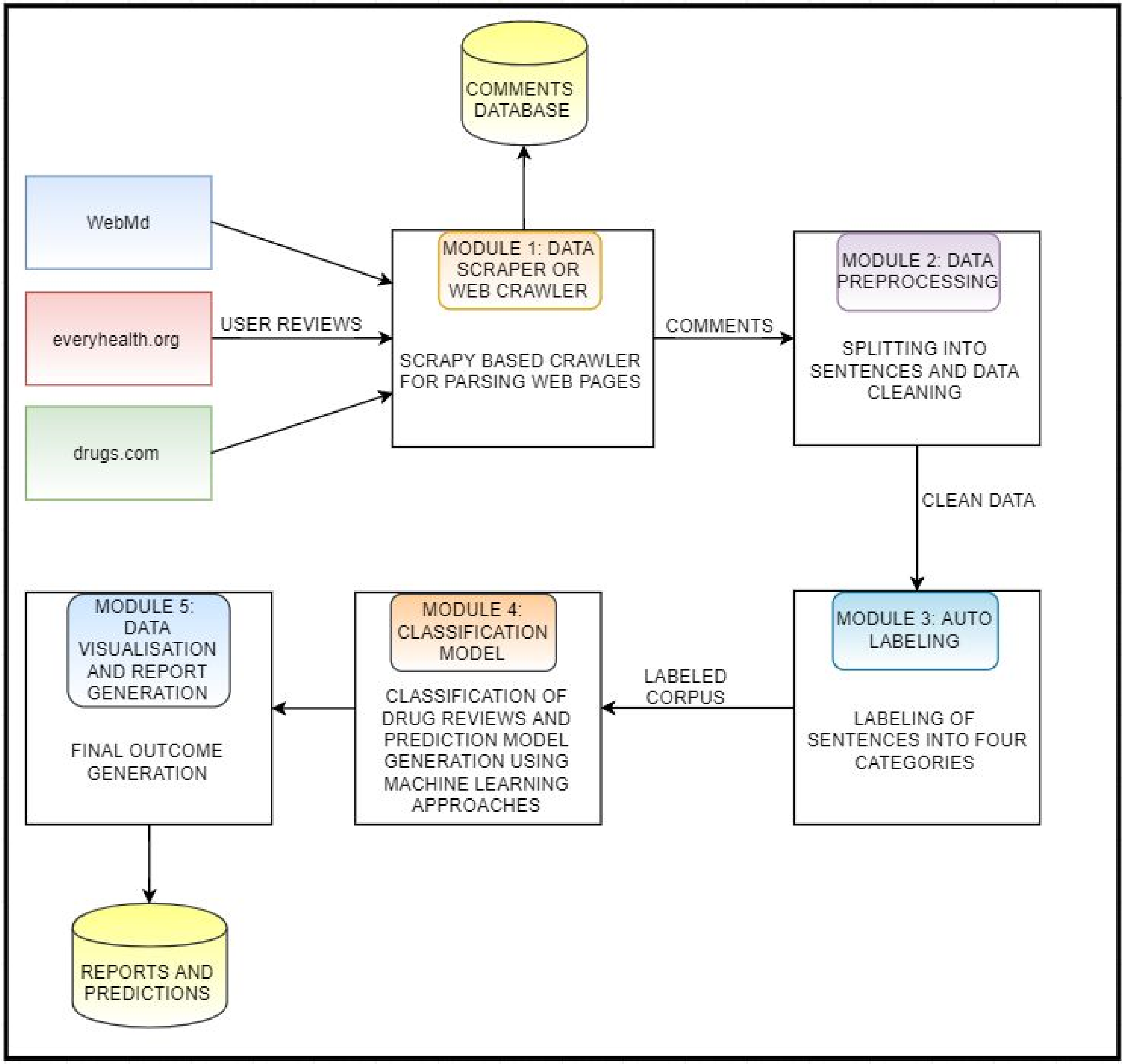


Fig 4.2 : Modular diagram of the developed system

Module 1: Data Scraper/ Web Crawler:

This module uses a scrapy based web crawler for parsing web pages. For this purpose user comments from three healthcare websites are scraped using a self-designed web crawler. The obtained comments are then stored in the database. We have scraped around 5000 user reviews from drugs.com, WebMd, everydayhealth.org. Scrapy based web crawler is smart enough to scrape most helpful comments and hence, crawler will crawl only scrape useful comments.

Module 2: Data Preprocessing:

The preprocessing module works on the obtained data. Preprocessing includes splitting the individual reviews into sentences, stemming and lemmatization. These steps provide clean data needed for training the model.

Module 3: Labeling of Data:

This phase includes labeling the clean data obtained from the preprocessing stage. The reviews are labeled based on four categories i.e. Effective, Ineffective, Adverse and Other. The reviews that lie in the 'Other' category are removed.This module provides data which is ready for training.

Module 4: Classification model:

The data is trained using a classification model. The model uses various machine learning approaches. The algorithms to be used include SVM, Random Forest, Logistic Regression, etc. The efficiency of each algorithm can be compared in order to obtain the most efficient algorithm.

Module 5: Visualization of data and Report generation:

Top ranked reviews of the drug under consideration are generated in this phase. A report including the pros and cons of the drug is also obtained. A comparison between different drugs for the same disorder can be performed. This could be used for suggestion of alternative drugs for a drug having severe adverse effects. Graphical analysis of various drugs with respect to a particular disease can also be obtained.

4.3 Detailed Design

a. DFD 0

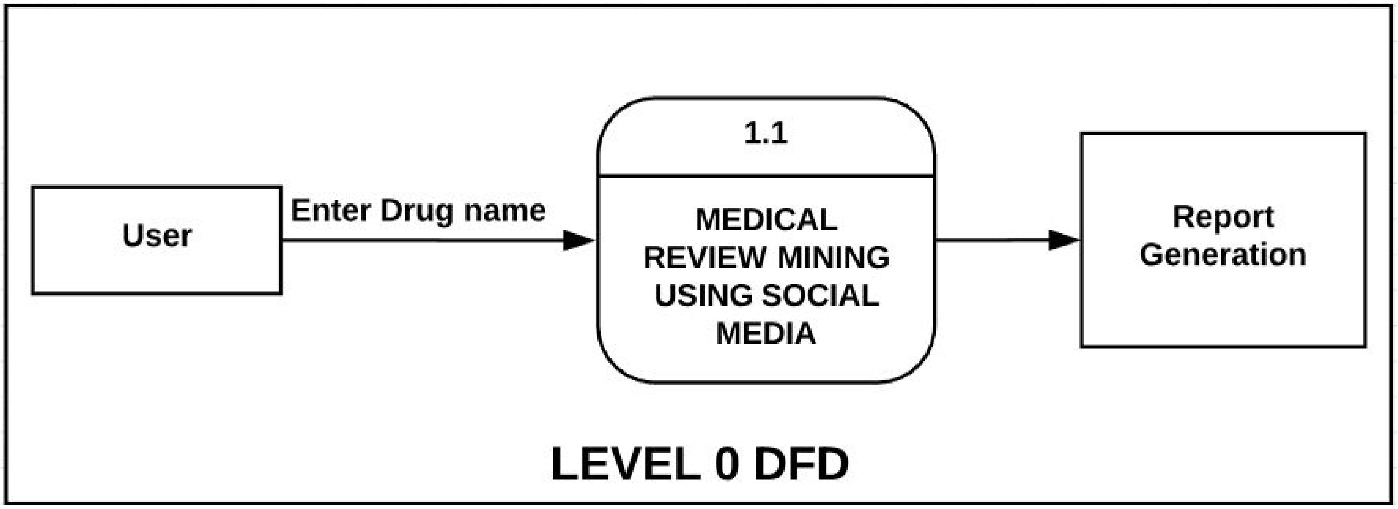


Fig 4.3.1: Level 0 DFD

The level 0 DFD shows the overall flow of the system. The user enters the name to the drug as an input to the system and the system displays top ranked reviews along with the pros and cons of the entered drug name.

b. DFD 1

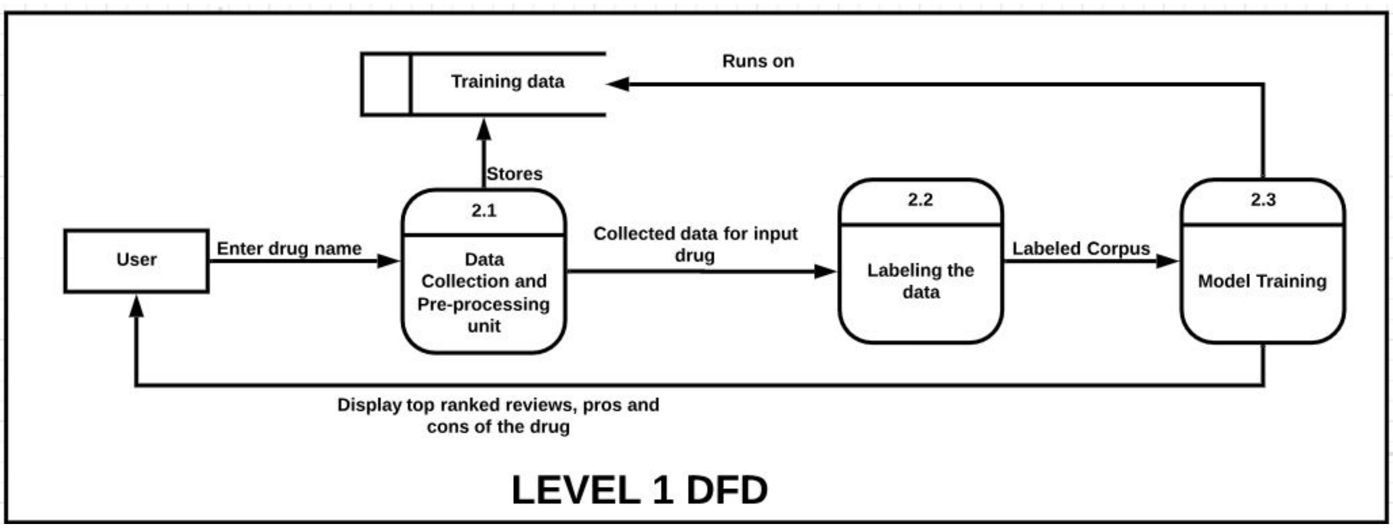


Fig 4.3.2 : Level 1 DFD

The level 1 DFD explains the various processes of the system in detail. The entered drug name is taken as an input by the data collection and preprocessing unit. This unit also stores the data. The data for the drug from the preprocessing unit is then taken for labeling. Once labeled, it is given as an input to the model training process. This process runs on the training data and displays the required output of the system.

c. DFD 2

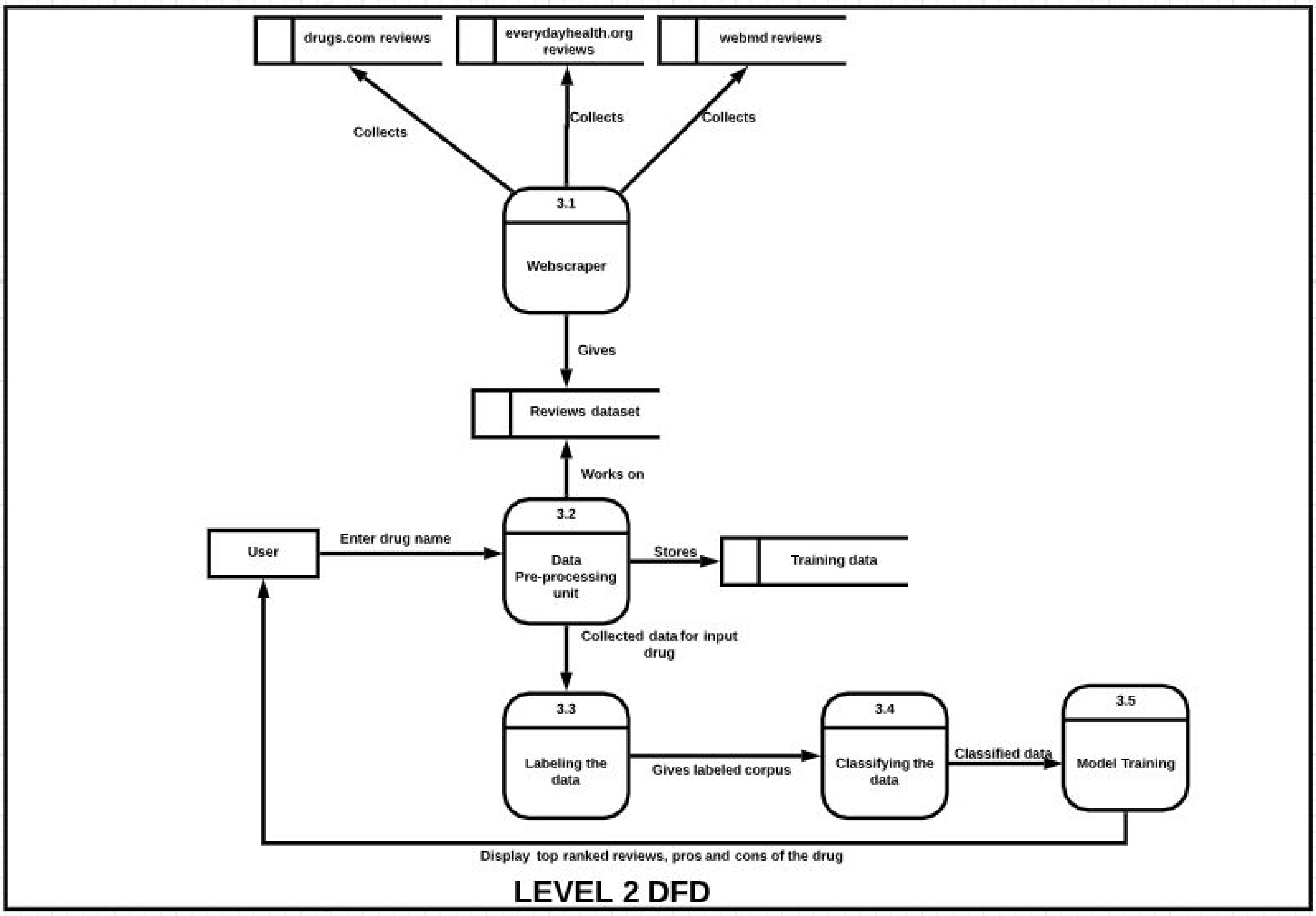
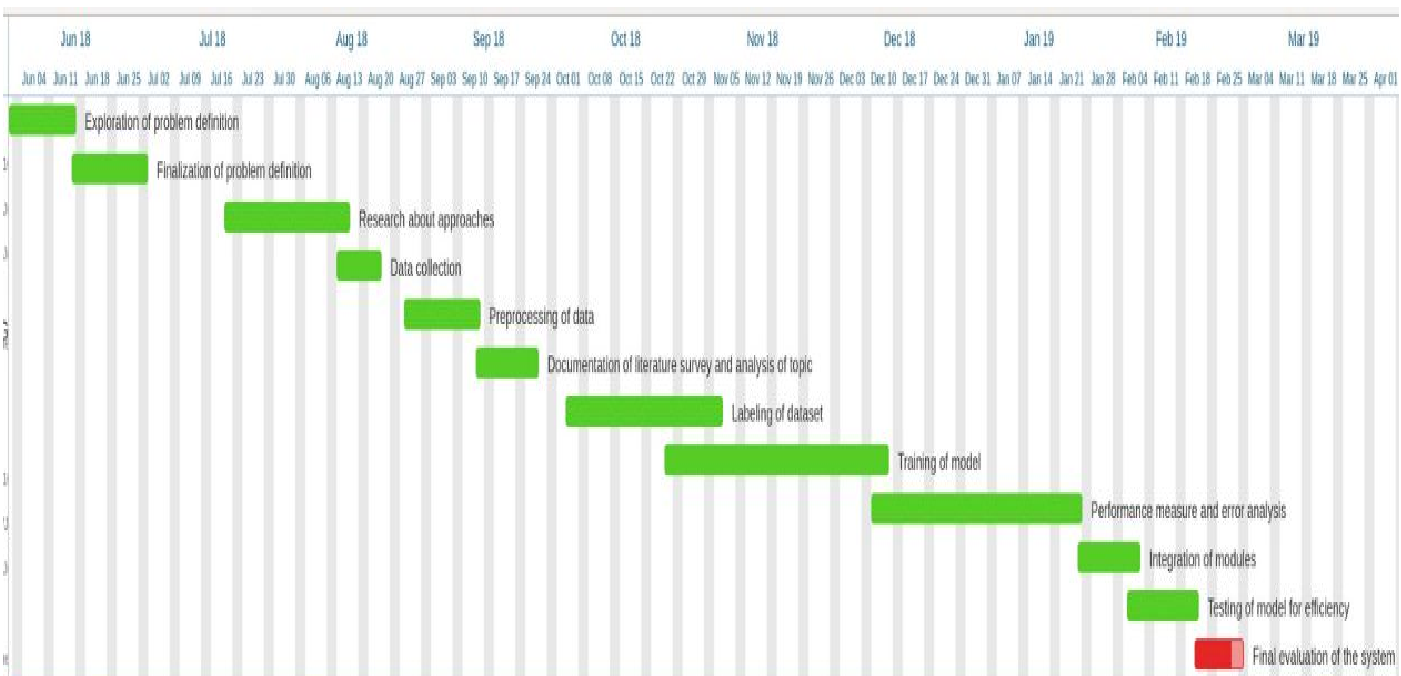


Fig 4.3.3 : Level 2 DFD

The level 2 DFD gives the detailed description of the system. The web scraper scrapes data from three websites, webmd, drugs.com and everydayhealth.org. The scraped data forms the reviews dataset. The preprocessing unit works on the reviews dataset for the user entered drug name. It also stores the training data. The preprocessed data is then labeled. The labeled corpus goes to the classification process. This process gives the output as classified data which is then given for model training. Finally, the output is displayed to the user.

4.4 Project Scheduling & Tracking using Time line / Gantt Chart

	Gantt	Sub Tasks	Assignee	Est. Hours	Start Date	Due Date	Task Prog.
0	Exploration of problem definition		Unassigned	7h	01/Jun	15/Jun	100%
1	Finalization of problem definition		Unassigned	8h	15/Jun	01/Jul	100%
2	Research about approaches		Unassigned	10h	19/Jul	15/Aug	100%
3	Data collection		Unassigned	15h	13/Aug	22/Aug	100%
4	Preprocessing of data		Unassigned	12h	28/Aug	13/Sep	100%
5	Documentation of literature survey and analysis of t…		Unassigned	5h	13/Sep	26/Sep	100%
6	Labeling of dataset		Unassigned	16h	03/Oct	06/Nov	100%
7	Training of model		Unassigned	15h	25/Oct	13/Dec	100%
8	Performance measure and error analysis		Unassigned	10h	10/Dec	25/Jan	100%
9	Integration of modules		Unassigned	24h	25/Jan	07/Feb	100%
10	Testing of model for efficiency		Unassigned	12h	05/Feb	20/Feb	100%
11	Final evaluation of the system		Unassigned	10h	20/Feb	02/Mar	75%



CHAPTER 5:

IMPLEMENTATION DETAILS

This chapter details the implementation details including the algorithms used, the methodology followed in other approaches following the results we obtained by our implementation.

5.1. Algorithms and flowcharts for the respective modules developed

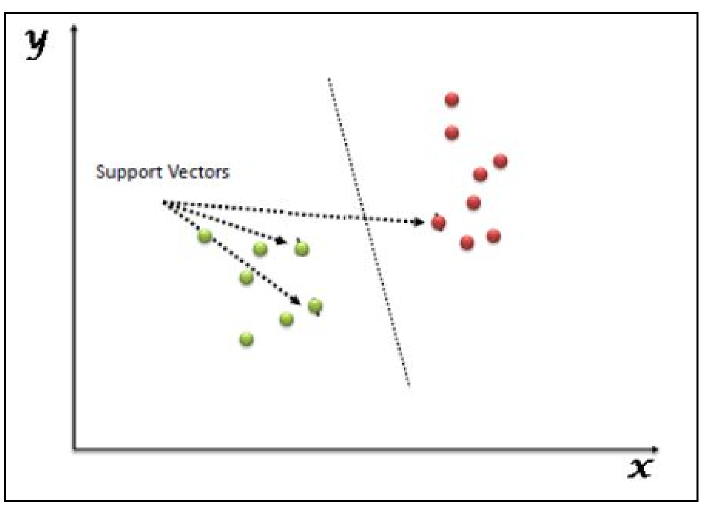
In our implementation, we have used OneVsRestClassifier, Also known as one-vs-all, this strategy consists of fitting one classifier per class. For each classifier, the class is fit against all the other classes. In addition to its computational efficiency (only n_classes classifiers are needed), one advantage of this approach is its interpretability. Since each class is represented by one and one classifier only, it is possible to gain knowledge about the class by inspecting its corresponding classifier. This is the most commonly used strategy for multiclass classification and is a fair default choice.

We have used this machine learning algorithm for implementing novel multi-class classifier. Multi-class classification means a classification task with more than two classes; each label is mutually exclusive. The classification makes the assumption that each sample is assigned to one and only one label.

Following are the learning algorithms:-

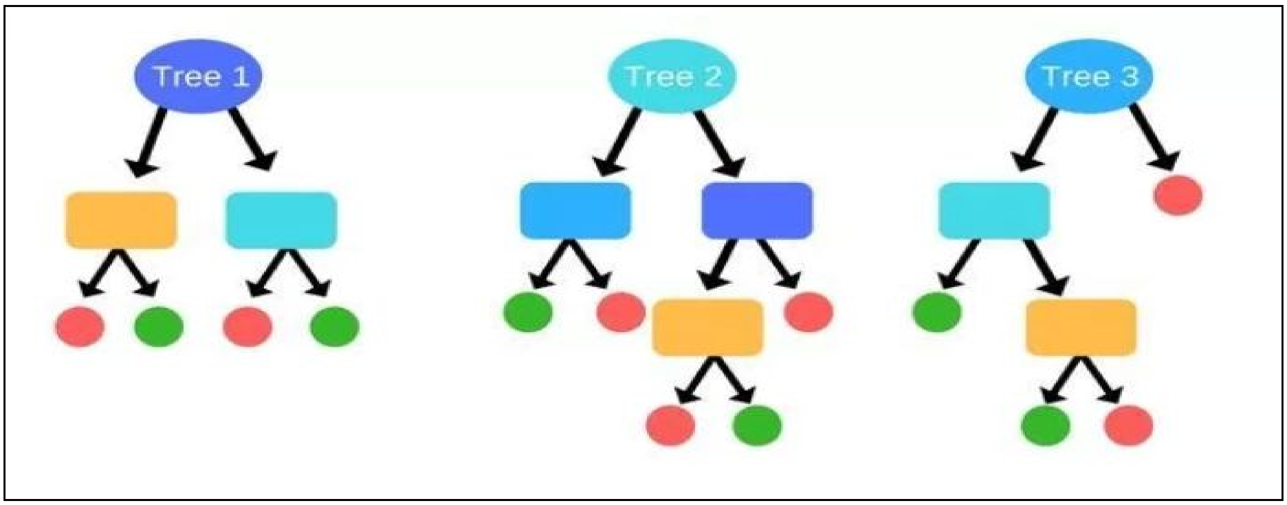
1. Support Vector Machine

   "Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification or regression problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyperplane that differentiate the two classes very well.



2. Random Forest Classification

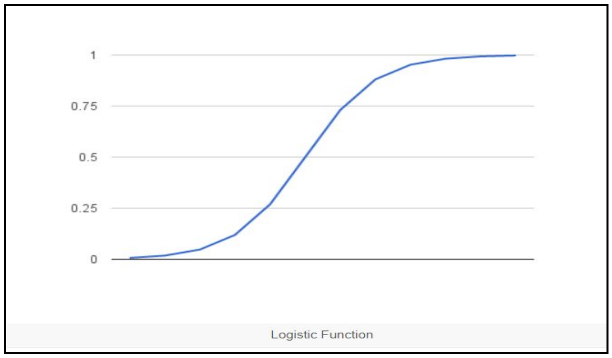
Random Forest is a supervised learning algorithm.it creates a forest and makes it somehow random.Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model. It is very easy to measure the relative importance of each feature on the prediction.
It randomly selects “k” features from total “m” features, where k << m. Among the “k” features, it calculates the node “d” using the best split point. It then splits the node into daughter nodes using the best split. The entire procedure is repeated until number of nodes has been achieved. There is no overfitting problem in classification using Random Forest Algorithm. Besides, The same random forest algorithm can be used for both classification and regression task.



3. Logistic Regression

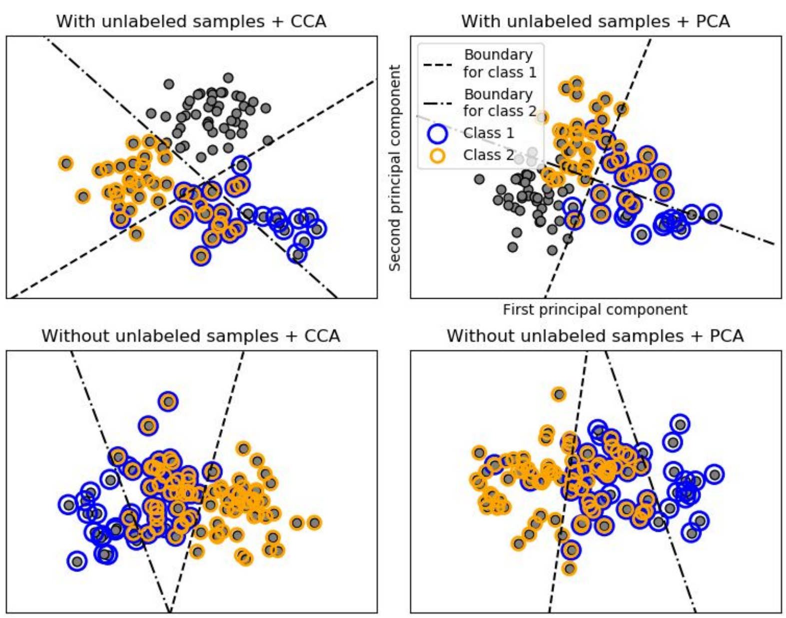
Logistic regression is another technique for classification problems. Logistic regression is named for the function used at the core of the method, the logistic function. The logistic function, also called the sigmoid function was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. It’s an S-shaped curve that can take any real-valued number and map it into a value between 0 and 1, but never exactly at those limits.

1 / (1 + e^-value)



4. OneVsRestClassifier

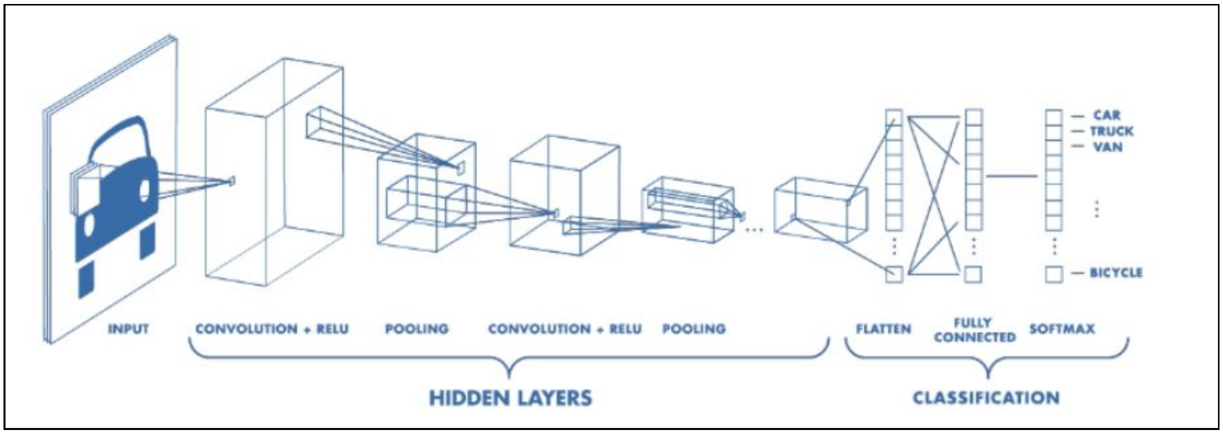
   Also known as one-vs-all, this strategy consists of fitting one classifier per class. For each classifier, the class is fitted against all the other classes. In addition to its computational efficiency (only n_classes classifiers are needed), one advantage of this approach is its interpretability. Since each class is represented by one and one classifier only, it is possible to gain knowledge about the class by inspecting its corresponding classifier. This is the most commonly used strategy for multiclass classification and is a fair default choice.



5. Convolution Neural Networks

   Convolutional Neural Networks or covnets are neural networks that share their parameters.

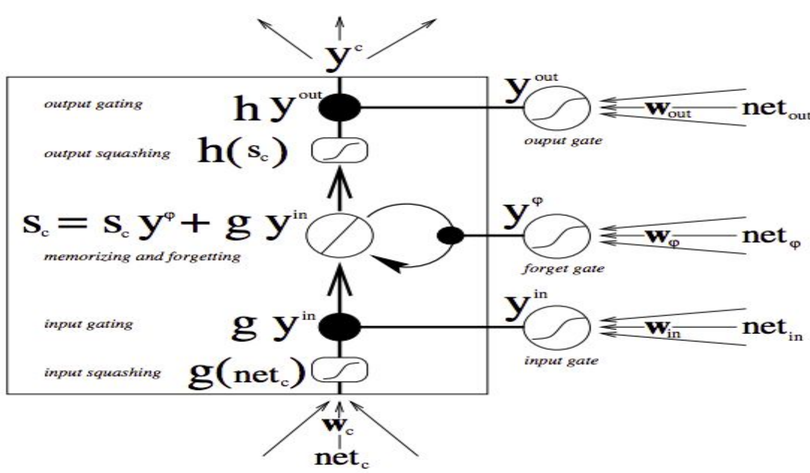
The Hidden layer performs a series of convolutions and pooling operations during which the features are detected. In the Classification part the fully connected layers will serve as a classifier on top of these extracted features. They will assign a probability for the object on the image being what the algorithm predicts it is. CNNs have high computational cost and need a lot of training data.



6. Long Short Term Memory

In the mid-90s, a variation of recurrent net with so-called Long Short-Term Memory units, or LSTMs, was proposed by the German researchers Sepp Hochreiter and Juergen Schmidhuber as a solution to the vanishing gradient problem.

LSTMs help preserve the error that can be backpropagated 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. This is one of the central challenges to machine learning and AI, since algorithms are frequently confronted by environments where reward signals are sparse and delayed, such as life itself.



Following are the modules of the implementation:-

1. Data Collection

   This module uses a scrapy based web crawler for parsing web pages. For this purpose user comments from three healthcare websites are scraped using a self-designed web crawler. The obtained comments are then stored in the database. We have scraped in all 3000 user reviews from drugs.com, WebMd, everydayhealth.org. Scrapy based web crawler is smart enough to scrape most helpful comments and hence, the crawler will crawl only scrape useful comments.

2. Data exploration

   Data exploration is very important when it comes to data mining. We have explored the data in various ways to extract our features.

3. Manual annotation

   This module includes labeling our dataset obtained from the crawler, only the descriptive reviews are used in this phase. Labeling is done on sentence level. The reviews are annotated into four categories i.e. Effective, Ineffective, Severe adverse reaction and Other. For this purpose, we make use of an annotation scheme for identifying the drug names and the relationships between the relevant keywords. Keywords indicate the category in which the review under consideration lies. This module provides dataset which is ready for pre-processing before training a classifier for prediction.

4. Pre-processing

   The reviews are pre-processed using gensim and python(3.6). It uses the combination of libraries such as nltk, gensim, spacy and basic python libraries. All the reviews are cleaned i.e removal of punctuations, conversion to lowercase and removal of all unuseful characters. Before considering the further cleaning, all reviews labelled as “OTHER” categories are discarded as they are creating noise in our dataset hence, reducing the efficiency.

   Following sub-modules for pre-processing

   a. Stopwords removal

Stopwords are a set of commonly used words in any language. The reason why stopwords are critical to many applications is that, if we remove the words that are very commonly used in a given language, we can focus on the important words instead.

b. Creation of bigrams and trigrams

c. Tokenization

A tokenizer divides a string into substrings by splitting on the specified string. We used natural language processing tool kit (NLTK) for tokenization.

5. Novel multi-class classifier

a. Deep learning approach

Long Short Term Memory model is designed for classification problems. Different variations of LSTM were used for obtaining accurate results. Also, we implemented combination of convolution neural network and LSTM model for efficient and precise results. The dataset used by us is limited with respect to the parameters considered. Hence, deep learning models were not as efficient due to problems such as imbalanced dataset which led to biased results and fewer samples for training and testing. The deep learning model achieved an accuracy of 80% on the dataset but is biased towards the dominant class, i.e effective. As a result, we preferred the machine learning approach for more accurate predictions.

b. Machine learning approach

We evaluated various classification algorithms of machine learning using scikit learn library. After considering all the precision reports we selected OneVsRestClassifier because of the accurate results obtained across all classes as compared to all other approaches we implemented. This classifier uses hybrid feature selection technique. TF-IDF, which stands for term frequency inverse document frequency, is a technique widely used in information retrieval (IR) or summarization problems. TF-IDF is intended to reflect how relevant a term is in a given document. We have used TF-IDF for feature creation for our classifier and sentiment scores obtained from VADER library as additional features.

6. Visualization and summarization:

Classified user reviews are considered for visualization of drug reports.

Following visualizations are considered:-

a. Pie Chart - Visualizing all the reviews for an individual drug in all categories.

b. Bar Chart - Dynamic barchart is implemented using bokeh library. Barchart gives the option to the user to choose drugs to plot for categories (Effective, Ineffective, Adverse).

c. LDA Graph- In addition to the machine learning model, we implemented LDA for extracting the words for each category. This graph is plotted using PyLDAvis library which takes an LDA model as input and gives clusters of words showing their percentage of occurrences in an individual category.

Summarization module gives top n reviews in each category. TextRank library is used for extracting the top n reviews. The user can view the top effective and top adverse reviews from extracted reviews.

Following is the code for the construction of OneVsRest classifier with Linear SVM core :

```
import numpy as np
import pandas as pd
from pprint import pprint
import keras


import codecs
import re
import seaborn as sns

import gensim
import gensim.corpora as corpora
from gensim.utils import simple_preprocess
from gensim.models import CoherenceModel
import spacy
import matplotlib.pyplot as plt

import warnings
warnings.filterwarnings("ignore",category=DeprecationWarning)
warnings.filterwarnings("ignore",category=FutureWarning)
```

```
reviews = pd.read_csv("All_Labelled_Reviews.csv")
reviews['category_id'] = reviews['Category'].factorize()[0]
reviews = reviews[reviews.category_id != 3]
reviews.reset_index(drop = True, inplace = True)
category_id_df = reviews[['Category',
'category_id']].drop_duplicates().sort_values('category_id')
id_to_category = dict(category_id_df[['category_id', 'Category']].values)
category_to_id = dict(category_id_df.values)
reviews.head()

from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer
def analyze_sentiment(df):
    sentiments = []
    sid = SentimentIntensityAnalyzer()
    for i in range(df.shape[0]):
        line = df['Review'].iloc[i]
        sentiment = sid.polarity_scores(line)
        sentiments.append([sentiment['neg'], sentiment['pos'],
                           sentiment['neu'], sentiment['compound']])
    df[['neg', 'pos', 'neu', 'compound']] = pd.DataFrame(sentiments)
#     df['Negative'] = df['compound'] < -0.1
#     df['Positive'] = df['compound'] > 0.1
    return df

analyze_sentiment(reviews).head()

from sklearn.model_selection import train_test_split

list_corpus = reviews["Review"].tolist()
list_labels = reviews["Category"].tolist()

import nltk
nltk.download('punkt')

from nltk.tokenize import word_tokenize
from sklearn.feature_extraction.text import CountVectorizer, TfidfVectorizer

vectorizer = TfidfVectorizer(ngram_range=(1, 2), tokenizer=word_tokenize,
max_features=15000)
vector = vectorizer.fit_transform(list_corpus)
```

```
df2 = pd.DataFrame(vector.toarray())
df2.head()

df_final = pd.concat([df2, reviews], axis =1)
df_final = df_final.drop('Review', axis =1)
df_final = df_final.drop('Category', axis =1)
df_final = df_final.drop('Drug', axis = 1)
df_final_labels = df_final['category_id']
df_final = df_final.drop('category_id', axis = 1)
df_final.head()

X_train, X_test, y_train, y_test,indices_train, indices_test =
train_test_split(df_final, df_final_labels,reviews.index, test_size=0.2,

random_state=42)

from sklearn.model_selection import cross_val_score
from sklearn.svm import LinearSVC
svc = LinearSVC(dual=False, multi_class='ovr', class_weight='balanced')
scores = cross_val_score(svc, X_train, y_train, scoring='f1_weighted', n_jobs=-1,
cv=10)
print('Cross-validation mean accuracy {0:.2f}%, std
{1:.2f}.'.format(np.mean(scores) * 100, np.std(scores) * 100))

svc.fit(X_train, y_train)
review = input("Enter your review:")
print(svc.predict([review].reshape(-1, 1)))
review = input("Enter your review:")

print(clf_tfidf_SVC.predict(tfidf_vectorizer.transform([review])))


pickle.dump(clf_tfidf_SVC, open('model.pkl','wb'))
pickle.dump(tfidf_vectorizer, open("vectorizer.pkl", "wb"))

# Loading model to compare the results
# model = pickle.load(open('model.pkl','rb'))
# print(model.predict(tfidf_vectorizer.transform([review])))
```

5.2. Comparative Analysis with the existing algorithms

Previously built system mostly deals with structured data i.e patient records from hospital, adverse reactions reports from pharma laboratories. All models focuses on adverseness of the drugs but do not consider the ineffectiveness of the drugs. Our system also addresses effectiveness of drugs.

Our system focuses on the user as the final outcome of the prediction should be easily comprehensible by the user in the form of simple visualisations and compact analysis.

CHAPTER 6:

TESTING

In this chapter, we have described the various types of testing which are performed on the system. The various test cases considered for testing and the inference drawn from those test cases are also described.

6.1. Definition of testing

Testing is a process of executing a program with the aim of finding the error. To make software perform well it should be error free. It is the process of evaluating a software item to detect differences between the given input and expected output. Testing assesses the quality of the product. Software testing is a process that should be done during the development process. In other words, software testing is a verification and validation process.

6.2. Types of tests

- Unit Testing- It focuses on the smallest unit of system design. It tests an individual unit or group of interrelated units. It is often done by the programmer by using sample input and observing its corresponding outputs.
- Integration Testing- The objective is to take unit tested components and build a program structure that has been dictated by design. Integration testing is testing in which a group of components is combined to produce output. Integration testing is of two types:
    a) Black Box Testing
    b) White Box Testing
- Regression Testing- Every time a new module is added leads to changes in the program. This type of testing makes sure that the whole component works properly even after adding components to the complete program.
- Alpha Testing- This is a type of validation testing. It is a type of acceptance testing which is done before the product is released to customers. It is typically done by QA professionals.
- Beta Testing-The beta test is conducted at one or more customer sites by the end-user of the software. This version is released for the limited number of users for testing in real-time environment.

- System Testing-In this software is tested such that it works fine for different operating systems. It is covered under the black box testing technique. In this, we just focus on the required input and output without focusing on internal working.
- Stress Testing-In this, unfavorable conditions are given to the system and check performance in those condition.

6.3. Type of Testing considered with justification

- **Unit testing**- All the units i.e model development, visualization, summary extraction etc. were tested individually before integration. This ensured that every unit functioned properly without any errors.
- **Integration testing**- Once all the units were developed they were integrated together and tested for any errors. All the individual units developed were integrated using flask framework.
- **System testing**- The project was tested for operating system compatibility. For this it was tested on both Windows and Linux(Ubuntu) operating systems. This ensured that a user having either of the two operating systems would be able to use the developed system without any difficulty.
- **Stress testing**- The system was stress tested by varying the number of reviews, from 100 to more than 3000, given for training the model. This would ensure smooth functioning of the system on varied inputs.

6.4 Various test case scenarios considered

- Using multiple operating systems for project demonstration viz. Windows(32 and 64 bit) and Ubuntu.
- Changing the number of reviews used for model training ranging from 100 to more than 3000.
- Checking for system compatibility with multiple browsers like Google Chrome, Firefox, IE, etc.
- Using reviews ranging from a single sentence to paragraph-long reviews for classifier prediction.

6.5. Inference drawn from the test

The model was found to predict the class of a review very accurately. This showed that the length of the review was not a constraint for prediction. It was also able to learn from thousands of reviews as training data. The system was found to be compatible with both the operating systems considered. It also displayed the exact same user interface on all the browsers considered.

CHAPTER 7 :

RESULT ANALYSIS

In this chapter, we have summarized all the results including User Interface and graphical outputs.

7.1. Modules under consideration

- Dynamicity of the model.
- Summarization of the reviews.

7.2. Parameters considered

Accuracy of the predictions

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

Sentiment scores of the reviews

	Review	Category	Drug	category_id	neg	pos	neu	compound
0	But since I started alternating I haven't had ...	Effective	Phenytoin	0	0.000	0.0	1.000	0.0000
1	About the Half Life of Phenytoin, about Sudden...	Adverse	Phenytoin	1	0.245	0.0	0.755	-0.5994
2	I think it's time to switch back to what works	Ineffective	Phenytoin	2	0.000	0.0	1.000	0.0000
3	SICK	Adverse	Phenytoin	1	1.000	0.0	0.000	-0.5106
4	It does work!"	Effective	Phenytoin	0	0.000	0.0	1.000	0.0000

Other model prediction metrics like f1-score, precision and recall

$$Precision = \frac{(TP)}{(TP + FP)}$$

$$Recall = \frac{(TP)}{(TP + FN)}$$

$$f1\ Score = 2 * (Recall * Precision) / (Recall + Precision)$$

We considered the weighted f1 score from sklearn for model evaluation to account for the class imbalance in the dataset, which takes into account the support (number of true instances) for each class and then finds their mean score.

$$F_\beta = (1+\beta^2)\cdot\frac{\text{precision}\cdot\text{recall}}{(\beta^2\cdot\text{precision})+\text{recall}}$$

7.3. Screenshots of User Interface (UI) for the respective module

1. Data collection

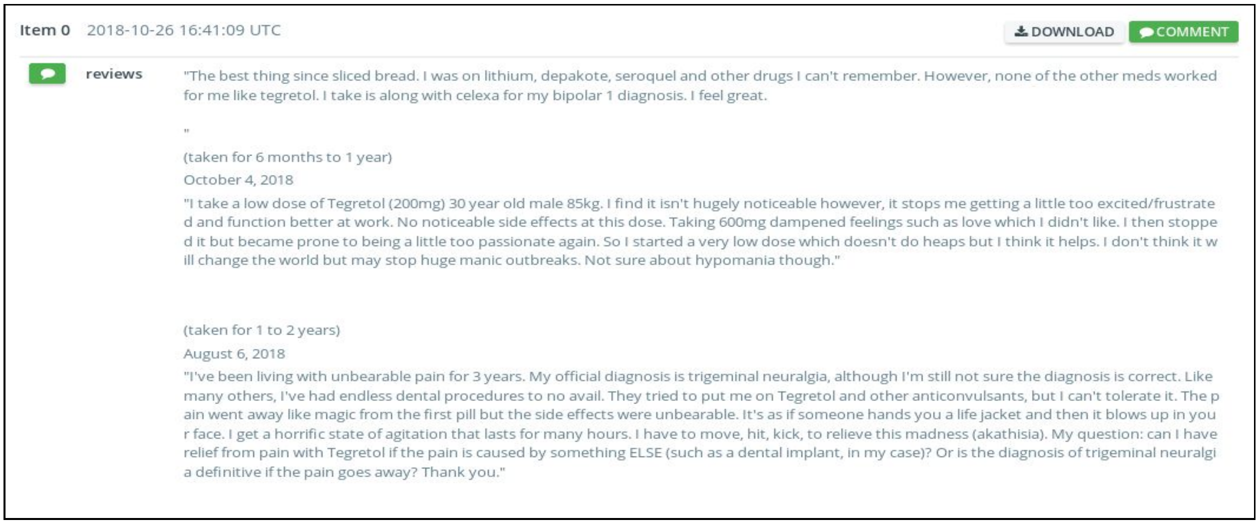


Figure 7.3.1: Scrapy based crawler for scraping all the reviews from websites.

2. Manual annotation of reviews

Review	Category	Drug
But since I started alternating I haven't had a seizure	Effective	Phenytoin
About the Half Life of Phenytoin, about Sudden unexpected death in epilepsy (SUDEP)	Adverse	Phenytoin
I think it's time to switch back to what works	Ineffective	Phenytoin
"Evening, I have been on this medication since I was 12 years of age	Other	Phenytoin
Why do doctors or neurologists NOT tell you about Epilepsy	Other	Phenytoin
Those who have Epilepsy do not necessarily know or at all ANYTHING about that cause	Other	Phenytoin
SICK	Adverse	Phenytoin
"I'm 24 years old and started to have seizures about 2 1/2 years ago	Other	Phenytoin
I had problems at first accepting that I was an epileptic and waited until i had my 5th seiz	Other	Phenytoin
I started on 300mg of phenytoin a day and went up to 500mg a day	Other	Phenytoin
First it was too little and then it was too much and really hurting my liver	Other	Phenytoin
Now I'm on 400-500mg a day alternating, which can be a pain in the butt trying to keep tra	Other	Phenytoin

Figure 7.3.2: Labelled Dataset

3. Exploratory Data Analysis

	Unnamed: 0	Review	Drug	category_id
Category				
Adverse	1089	1089	1089	1089
Effective	1276	1276	1276	1276
Ineffective	335	335	335	335

Figure 7.3.3: Number of reviews in each category

	Category	word_count
0	Adverse	16.496786
1	Effective	15.543887
2	Ineffective	14.504478

Figure 7.3.4: Word count in each category

	comments	word_count
0	"I am now 78 yrs old. Have taken 15mg in AM, ...	134
1	"I have been taking it for 15 years and no maj...	39
2	"Took care of it after a couple days."	8
3	"It was ok at first but the second day I could...	13
4	"Taken this medicine for decades and so far no...	28

Figure 7.3.5: Word count for each review

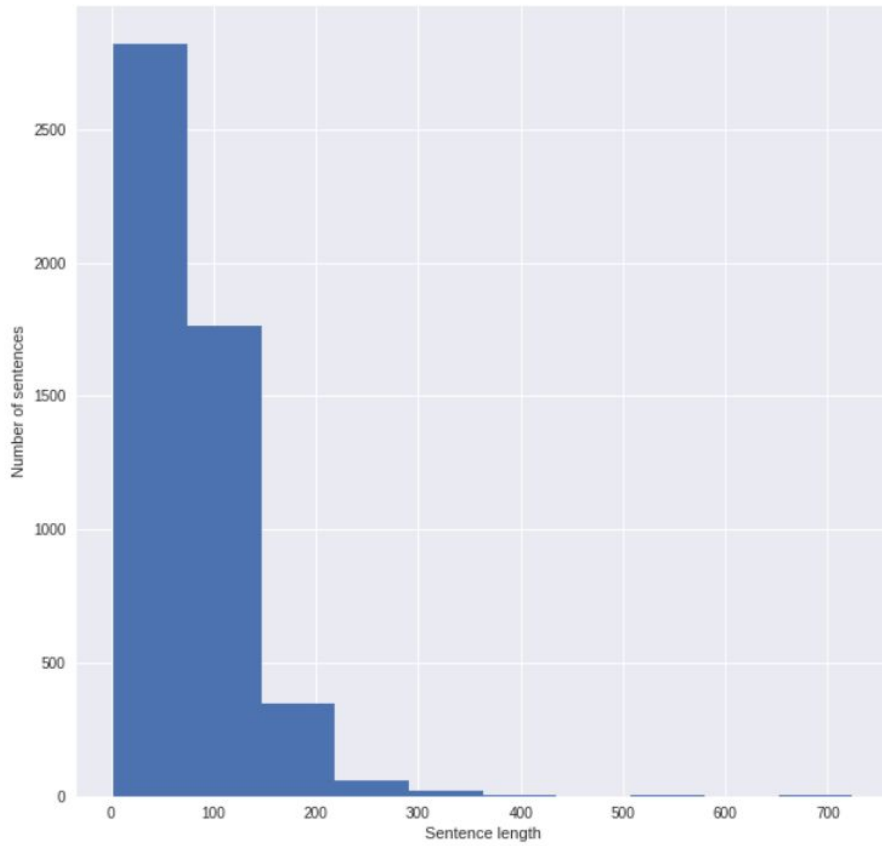


Figure 7.3.6: Maximum sentence length

	comments	stopwords
0	"I am now 78 yrs old. Have taken 15mg in AM, ...	33
1	"I have been taking it for 15 years and no maj...	16
2	"Took care of it after a couple days."	4
3	"It was ok at first but the second day I could...	4
4	"Taken this medicine for decades and so far no...	13

Figure 7.3.7: Stopwords in each review

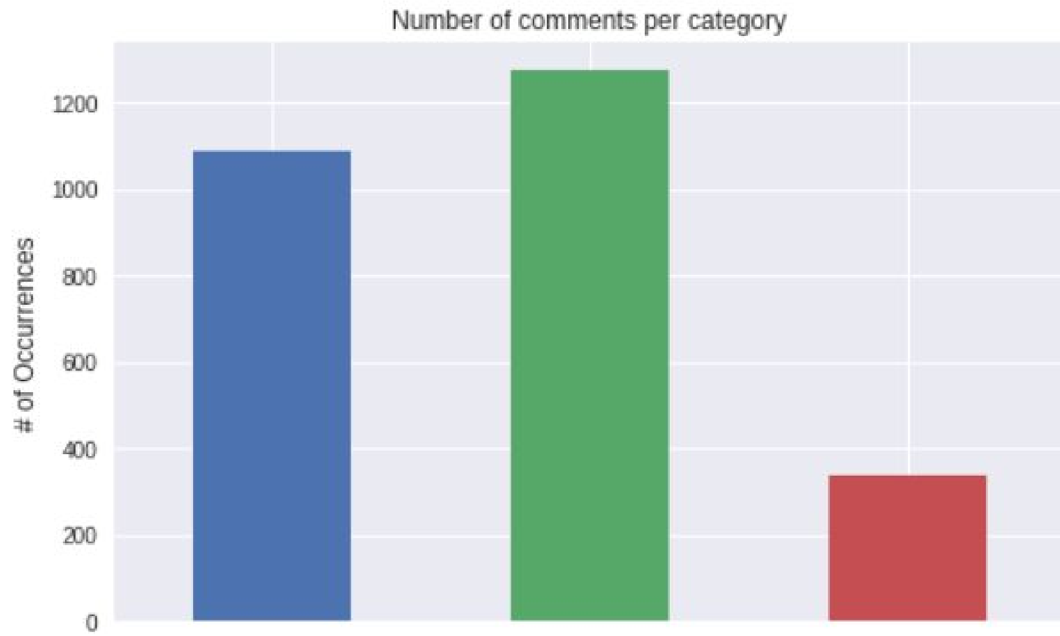


Figure 7.3.8: Reviews per category

4. Pre-processing

Before pre-processing:-

"But since I started alternating, I haven't had a seizure"

After pre-processing:-

```
[['but', 'since', 'started', 'alternating', 'haven', 'had', 'seizure']]
```

5. Multiclass classifier

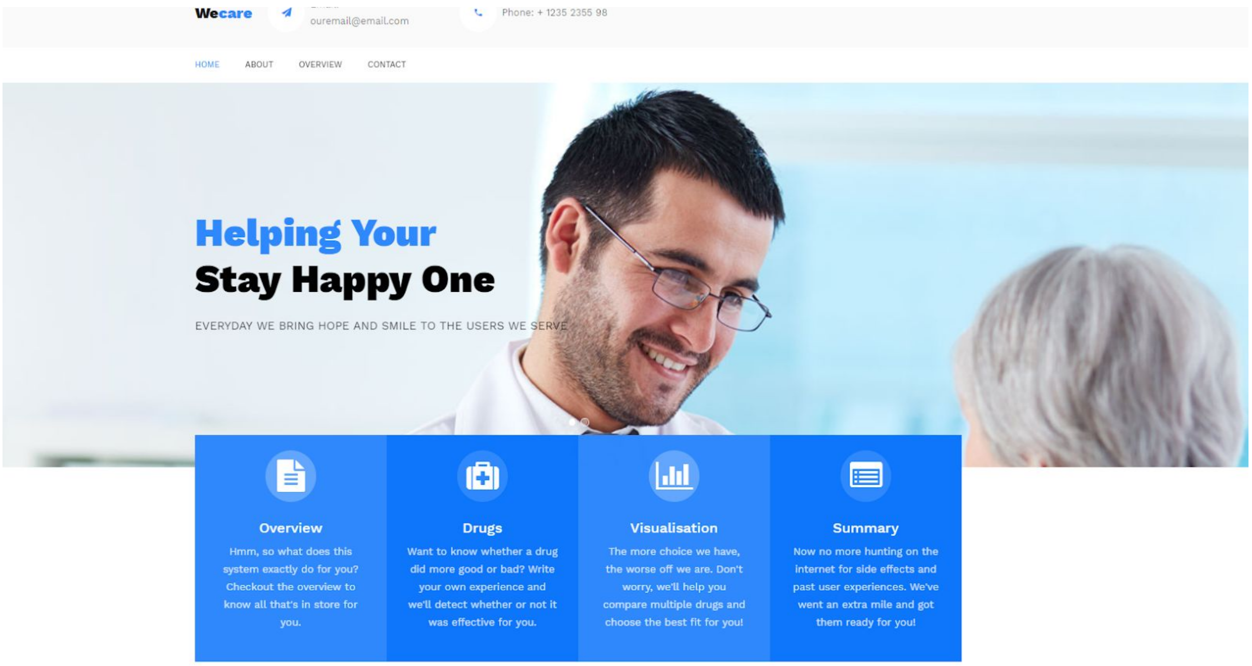


Figure 7.3.9: Main Home page for user

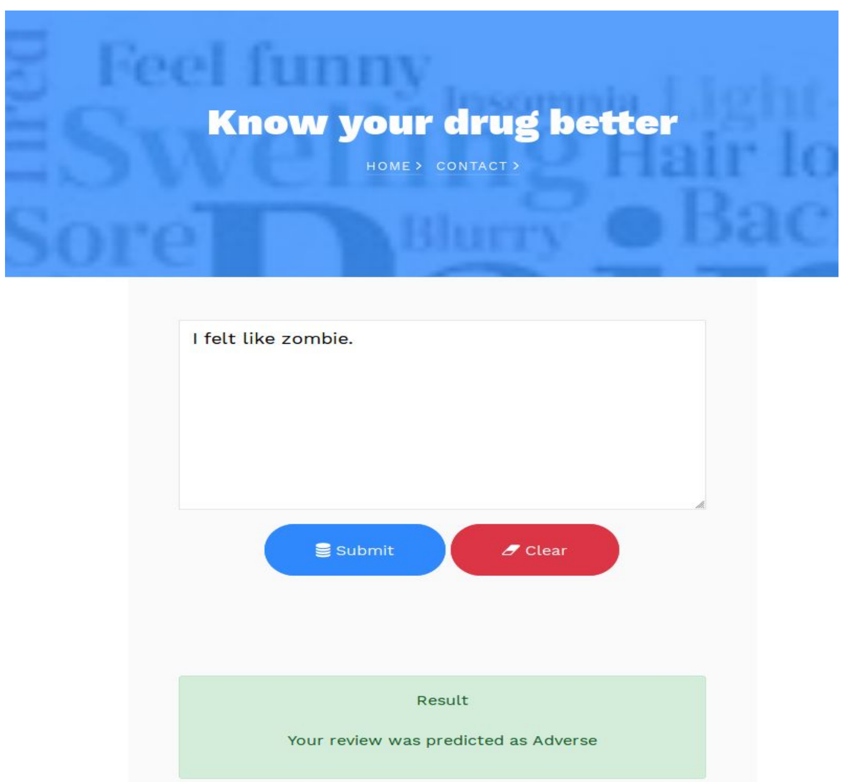


Figure 7.3.10: Novel Multi-Class prediction model

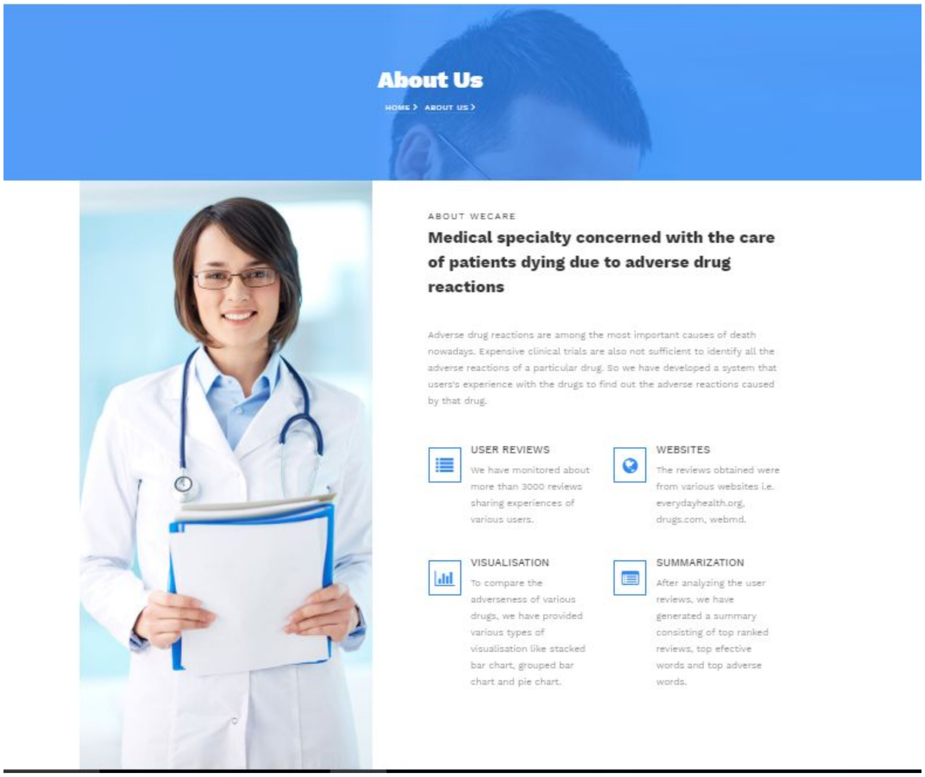


Figure 7.3.11: About Us

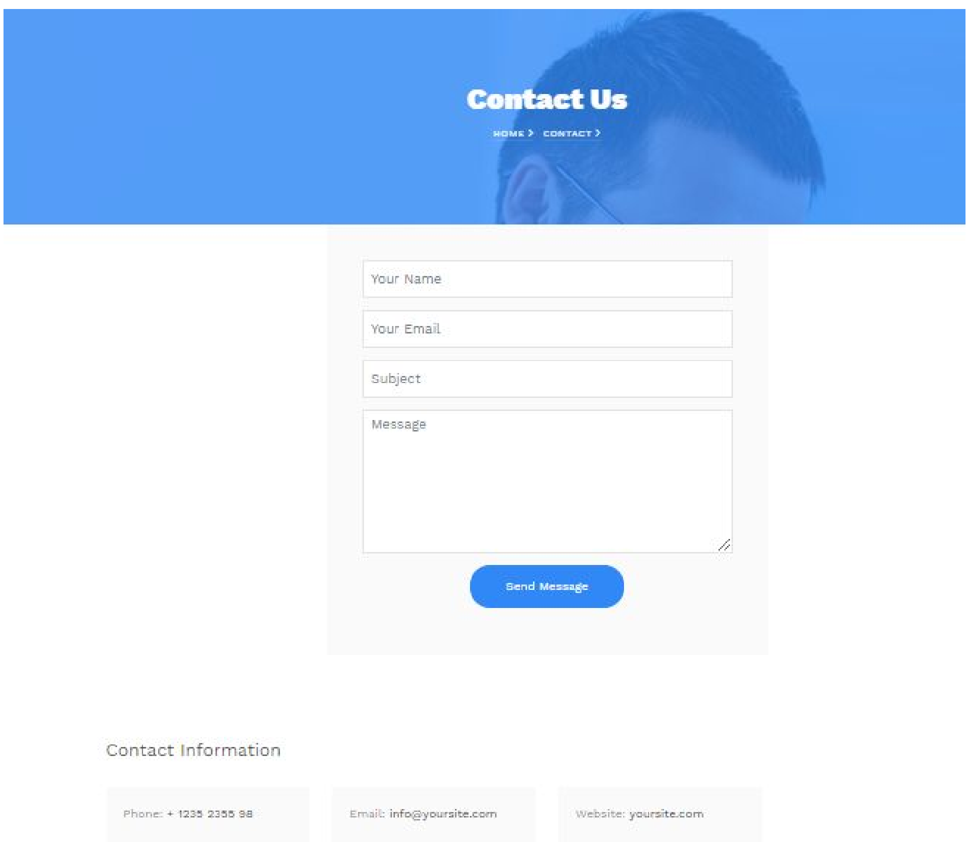


Figure 7.3.12: Contact Us

6. Summarization

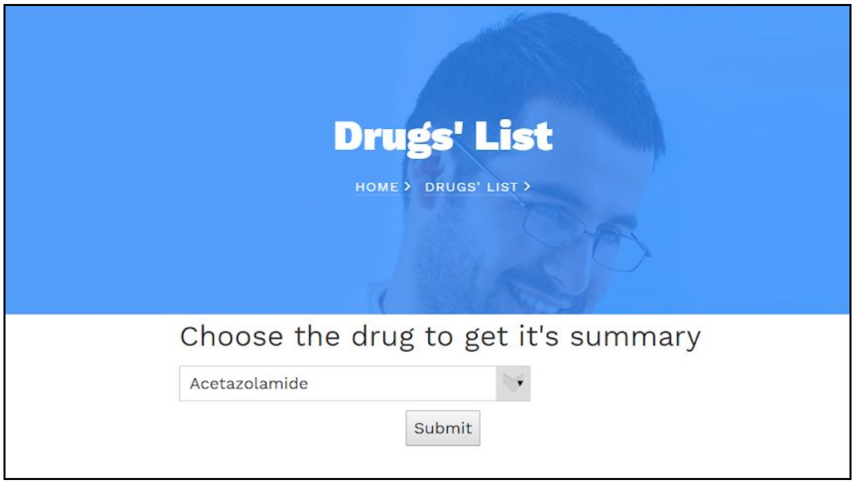
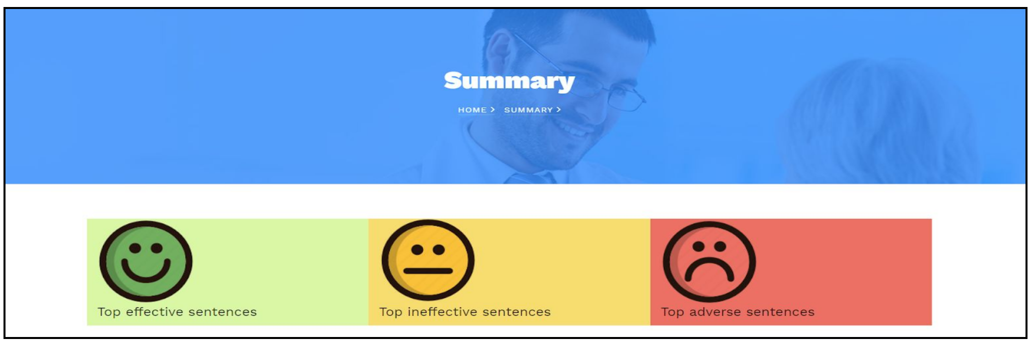
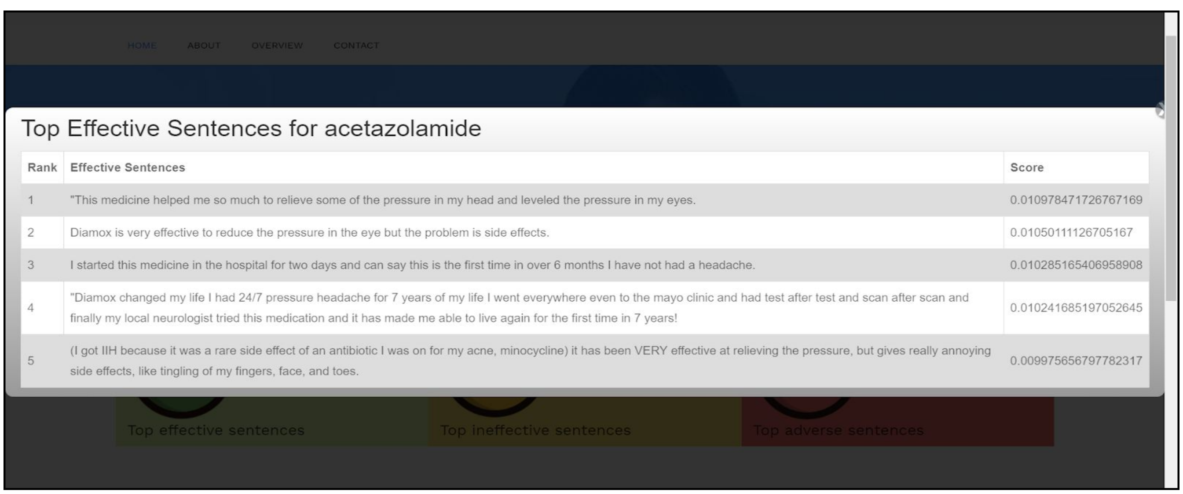


Figure 7.3.13: UI for list of drugs to choose from





Rank	Effective Sentences	Score
1	"This medicine helped me so much to relieve some of the pressure in my head and leveled the pressure in my eyes.	0.010978471726767169
2	Diamox is very effective to reduce the pressure in the eye but the problem is side effects.	0.01050111126705167
3	I started this medicine in the hospital for two days and can say this is the first time in over 6 months I have not had a headache.	0.010285165406958908
4	"Diamox changed my life I had 24/7 pressure headache for 7 years of my life I went everywhere even to the mayo clinic and had test after test and scan after scan and finally my local neurologist tried this medication and it has made me able to live again for the first time in 7 years!	0.010241685197052645
5	(I got IIH because it was a rare side effect of an antibiotic I was on for my acne, minocycline) it has been VERY effective at relieving the pressure, but gives really annoying side effects, like tingling of my fingers, face, and toes.	0.009975656797782317

Figure 7.3.14: UI for top effective sentences

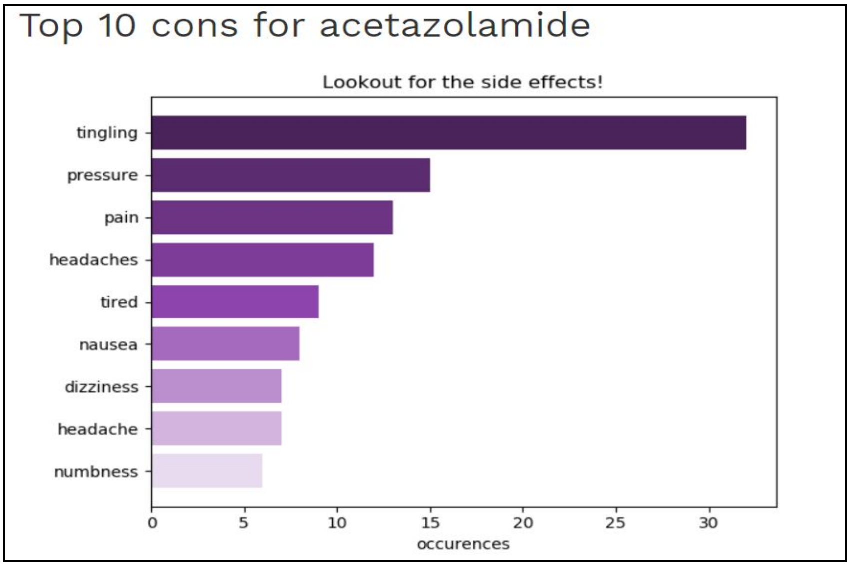
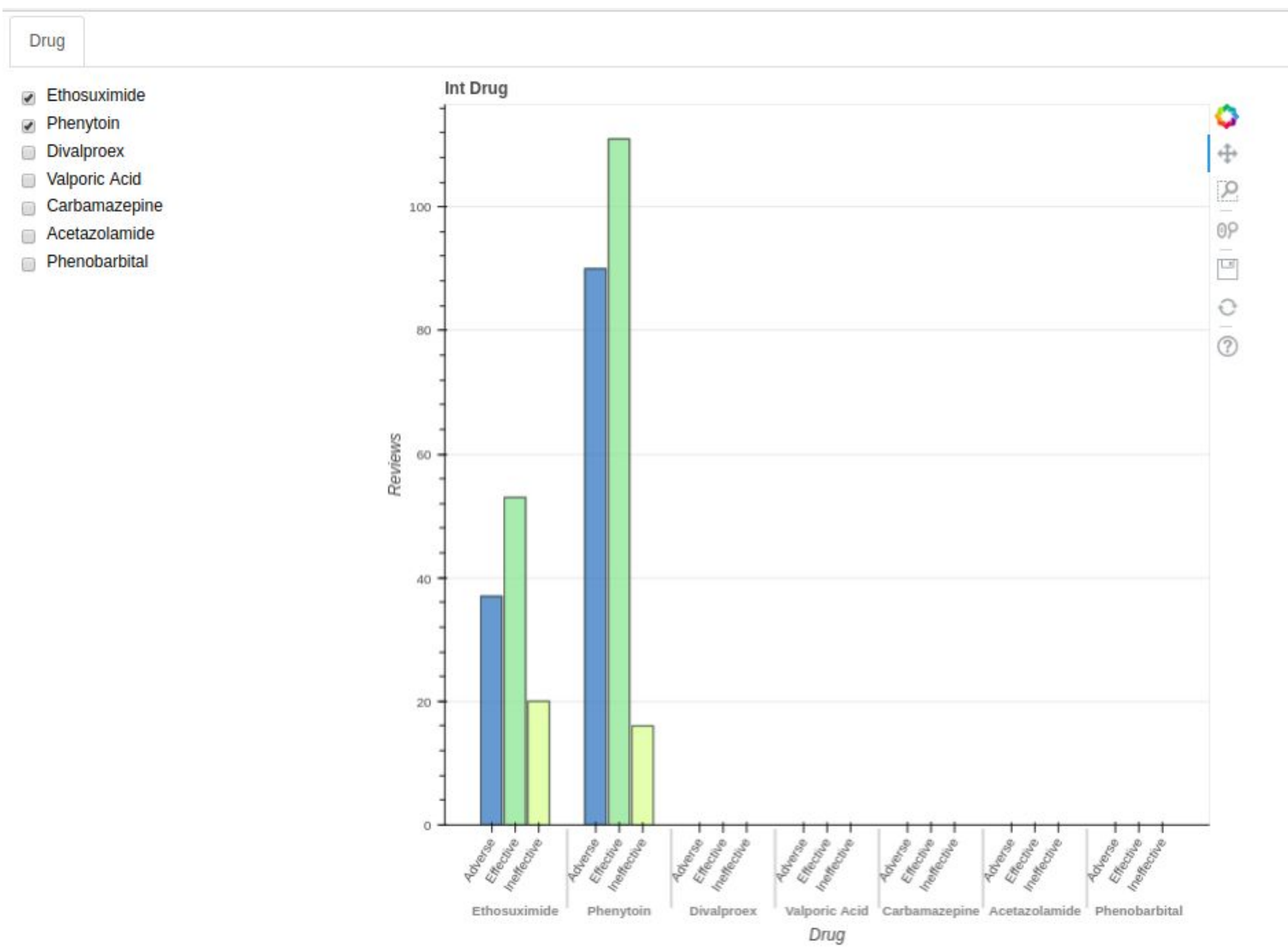


Figure 7.3.15: UI for top 10 cons

7.4. Graphical outputs of the various scenarios considered



..

Figure 7.4.1 Bar chart for two drugs selected (Ethosuximide and Phenytoin)

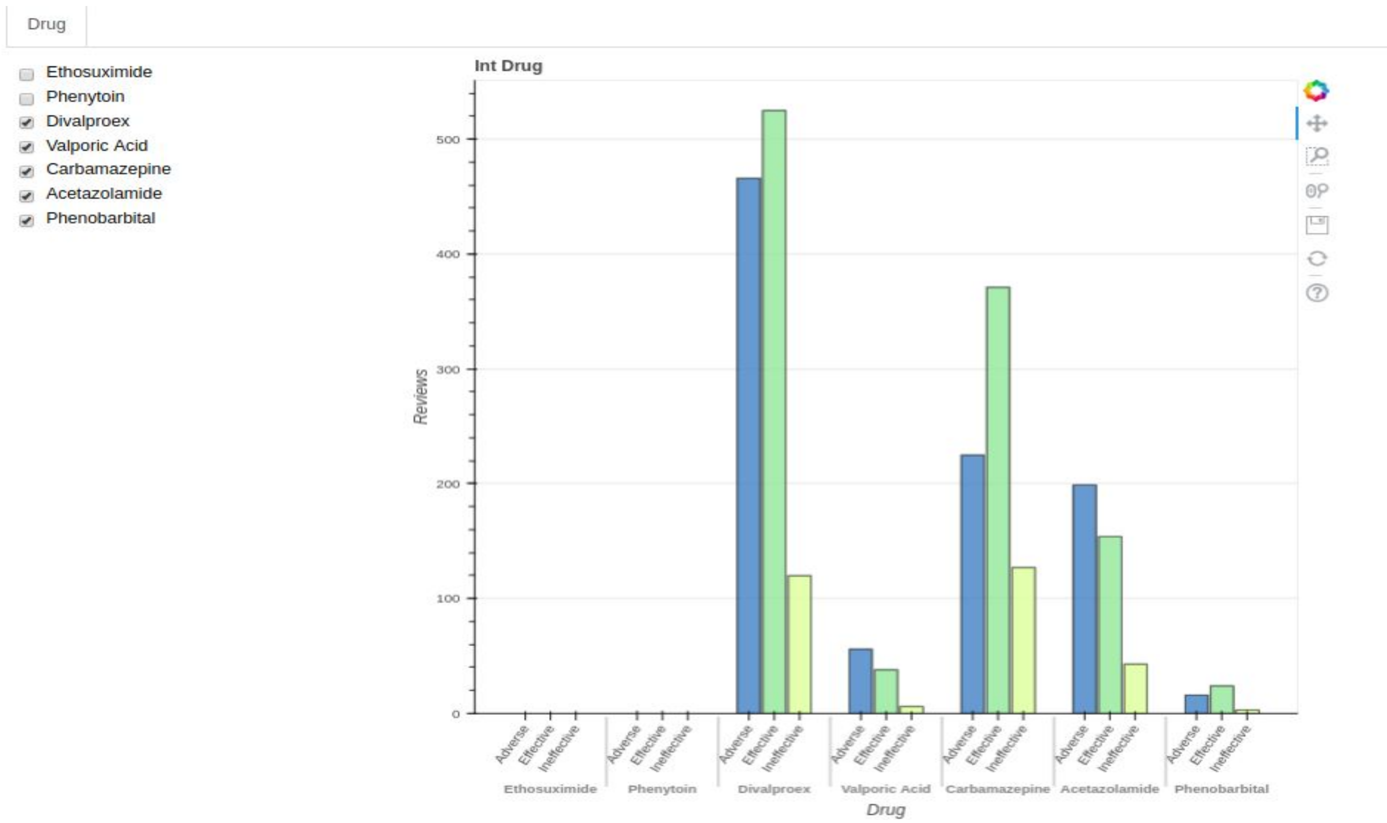


Figure 7.4.2 Bar chart for Ethosuximide and Phenytoin deselected

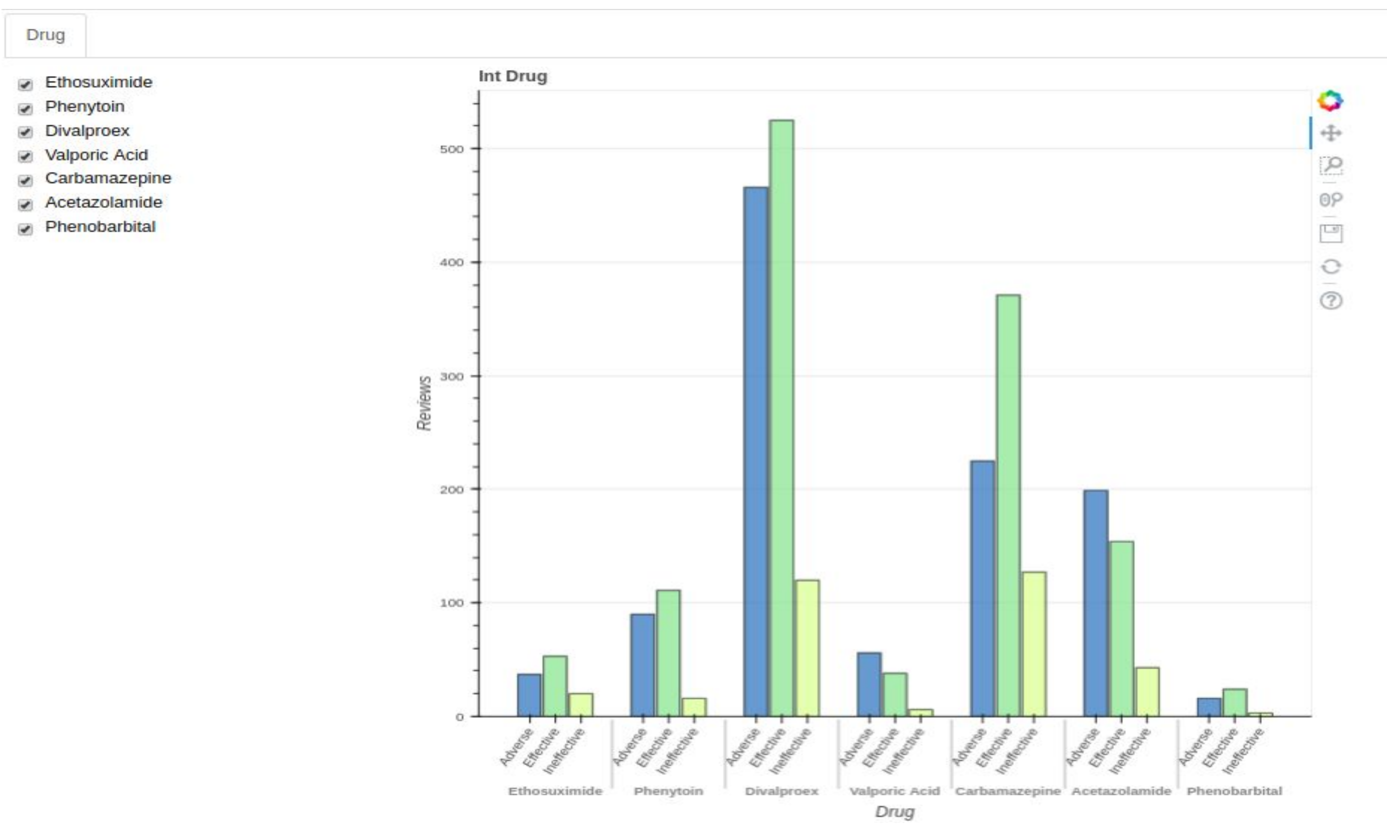


Figure 7.4.3 Bar chart for all drugs selected

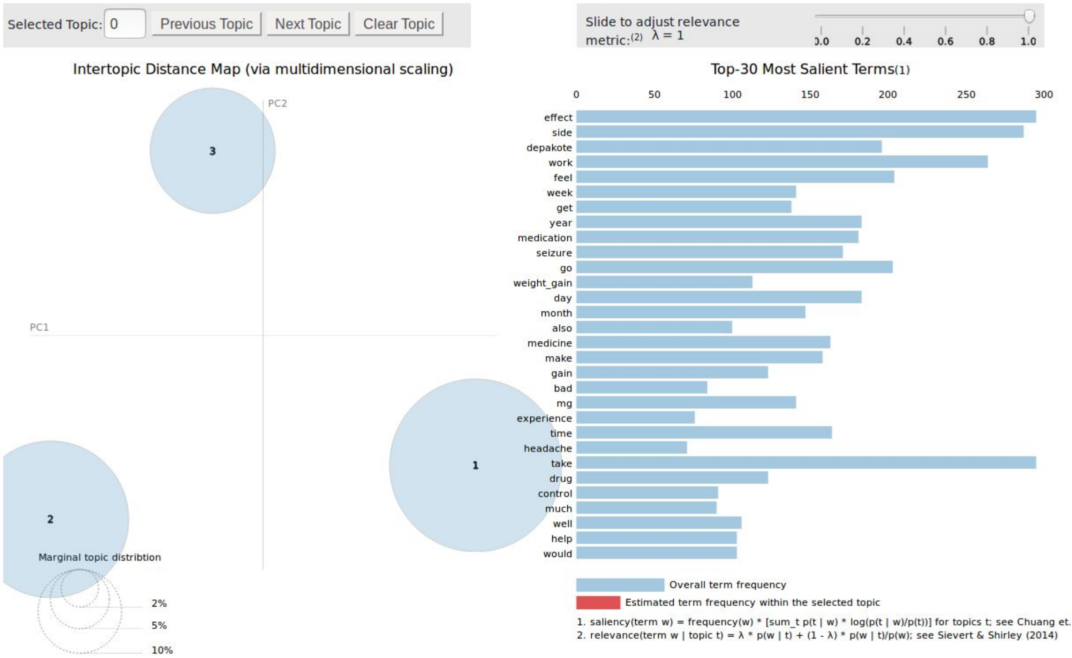


Figure 7.4.4 LDA model for reviews

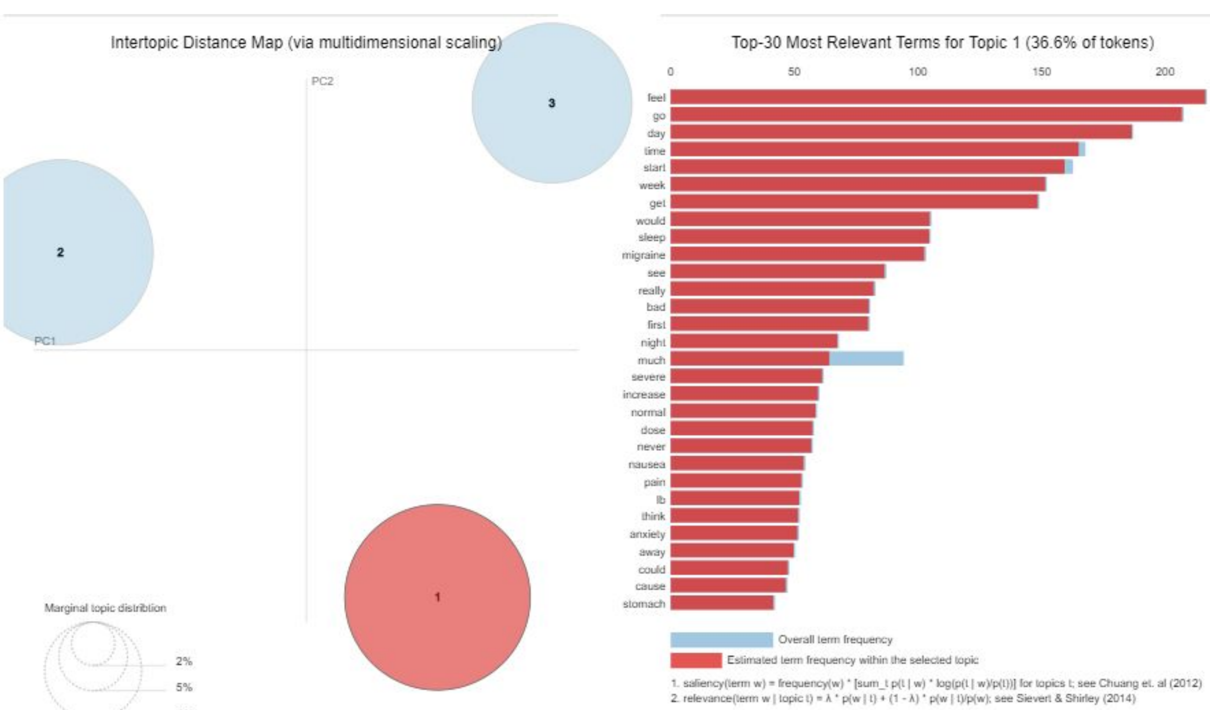


Figure 7.4.5 LDA model for Adverse reviews

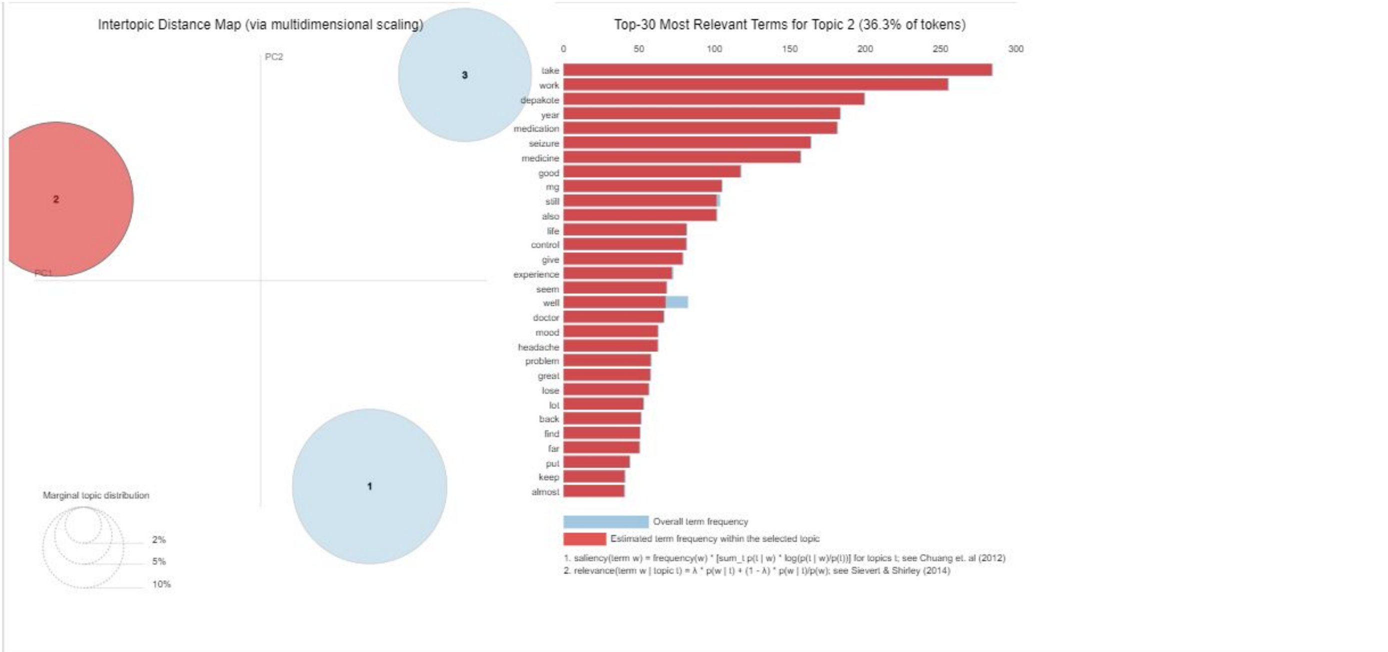


Figure 7.4.6 LDA model for Effective reviews

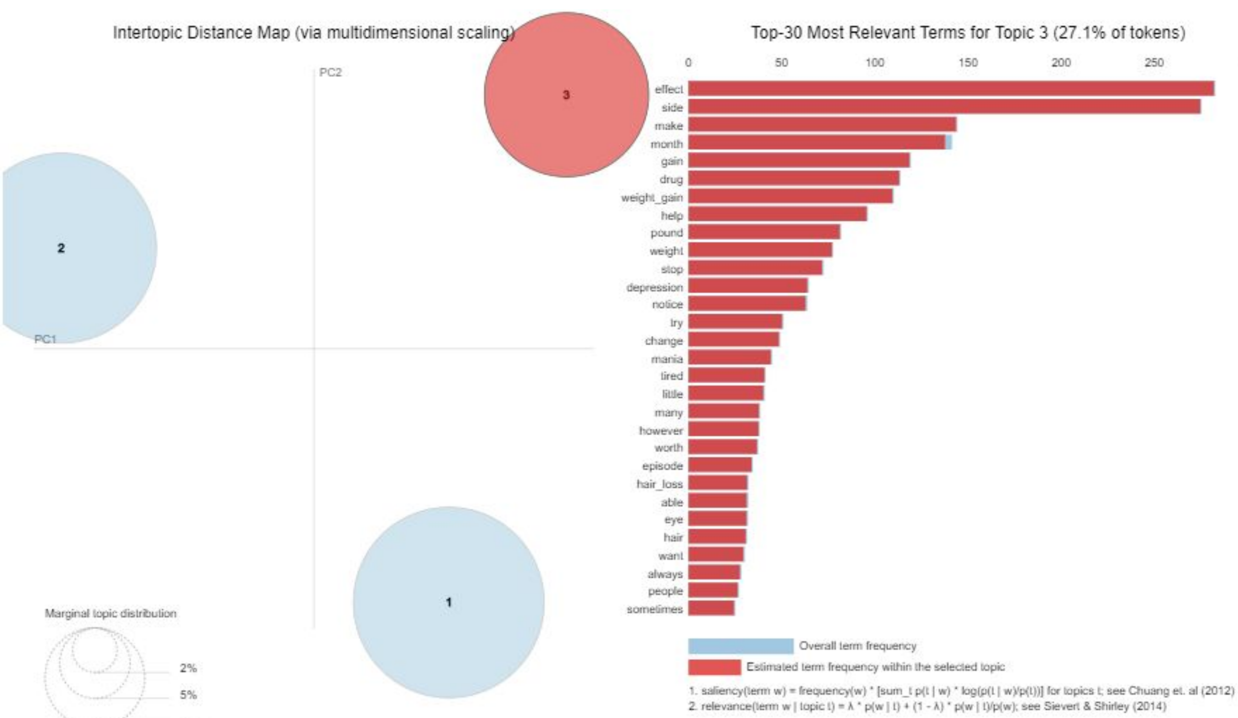


Figure 7.4.7 LDA model for Ineffective reviews

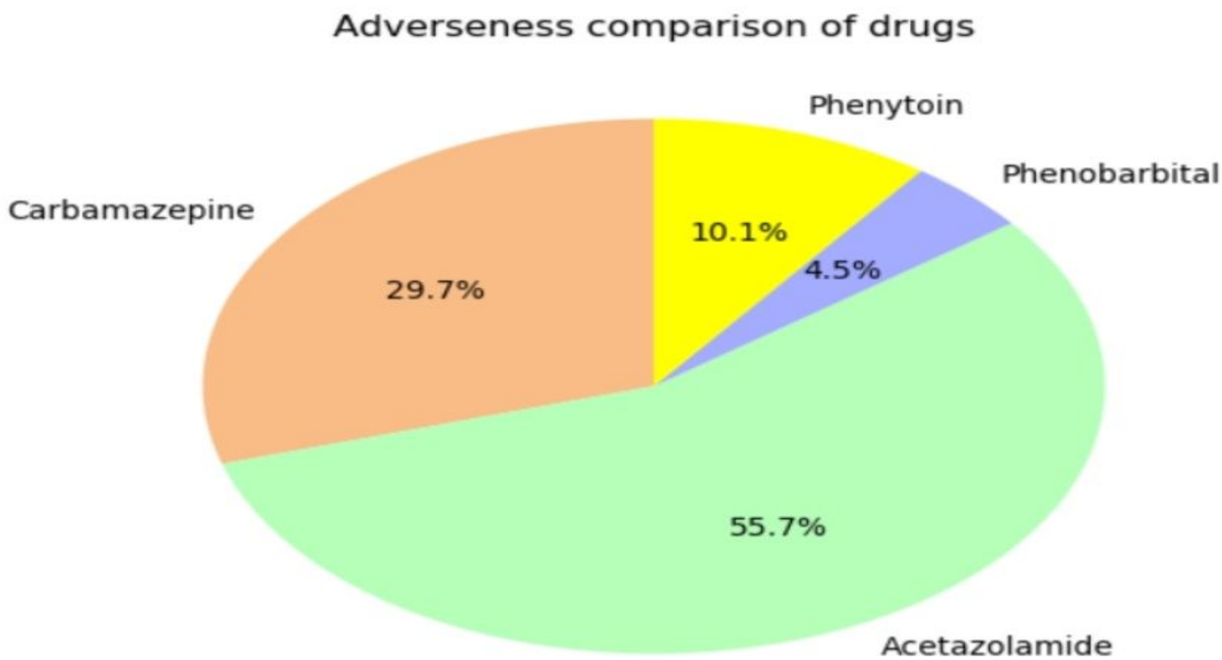


Figure 7.4.8 Pie Chart to compare adverseness of drugs

7.5. Evaluation of the developed system

The highest **accuracy** of the model is obtained by feeding a combination of top 15000 TF-IDF features from the review texts and VADER sentiment scores to a OnevsRestClassifier with a Linear SVM core and cross-validated across 10 folds. This results in an accuracy of ~75% across three categories - adverse, effective, ineffective. The detailed output is shown in table 7.6.1. Accuracies obtained by other algorithms are shown in table 7.6.2

The **effectiveness** of our system is evident by the fact that our system that multiple classifiers and approaches are used for classification of reviews and it correctly classifies the majority of sentences provided during the testing period.

Our system caters to a wide range of user needs like visualisation and text analysis of reviews with summarisation in a simple yet slick manner. This shows the **efficiency** of the system with respect to user and professional needs.

7.6. Reports generated / Tables obtained

	precision	recall	f1-score	support
Effective	0.76	0.81	0.79	250
Adverse	0.75	0.74	0.75	232
Ineffective	0.42	0.33	0.37	58
micro avg	0.73	0.73	0.73	540
macro avg	0.65	0.63	0.63	540
weighted avg	0.72	0.73	0.72	540

Table 7.6.1 - Detailed analysis of OneVsRest Classifier.

We consider the OneVsRest Classifier with a linear SVM core as our final classifier as our final model because it is efficient across all classes and is not biased towards one single class, as shown in table 7.6.1.

Here, we show the comparison of all approaches with different feature selections in table 7.6.2

FS-1 : CountVectorizer

FS-2 : CountVectorizer + VADER Sentiment Scores

FS-3 : CountVectorizer top 10000 features + VADER Sentiment Scores + n-gram range 1-3

FS-4 : CountVectorizer all features + VADER Sentiment Scores + n-gram range 1-3

FS-5 : TfidfVectorizer

FS-6 : TfidfVectorizer + VADER Sentiment Scores

FS-7 : Tfidf Vectorizer top 10000 features + VADER Sentiment Scores + n-gram range 1-3

FS-8 : Tfidf Vectorizer top 15000 features + word tokenize analyser + VADER Sentiment Scores + n-gram range 1-3

	FS-1	FS-2	FS-3	FS-4	FS-5	FS-6	FS-7	FS-8
OneVsRest Classifier with linear SVM core	67.21	69.21	70.42	72.07	69.83	71.11	72.61	73.37
Logistic Regression	67.11	68.48	70.80	72.25	68.43	70.32	72.71	73.56
Random Forest	66.30	67.93	67.58	66.35	64.54	66.70	65.84	68.10
Bagging meta-estimator with linear regressor core	67.21	69.12	71.96	-	69.38	71.37	71.82	72.77

Table 7.6.2 Various approaches, feature selections and their respective weighted f1-scores

All scores printed are weighted f1 scores averaged for each of the three categories - effective, adverse, ineffective - and cross validated over 10 folds.

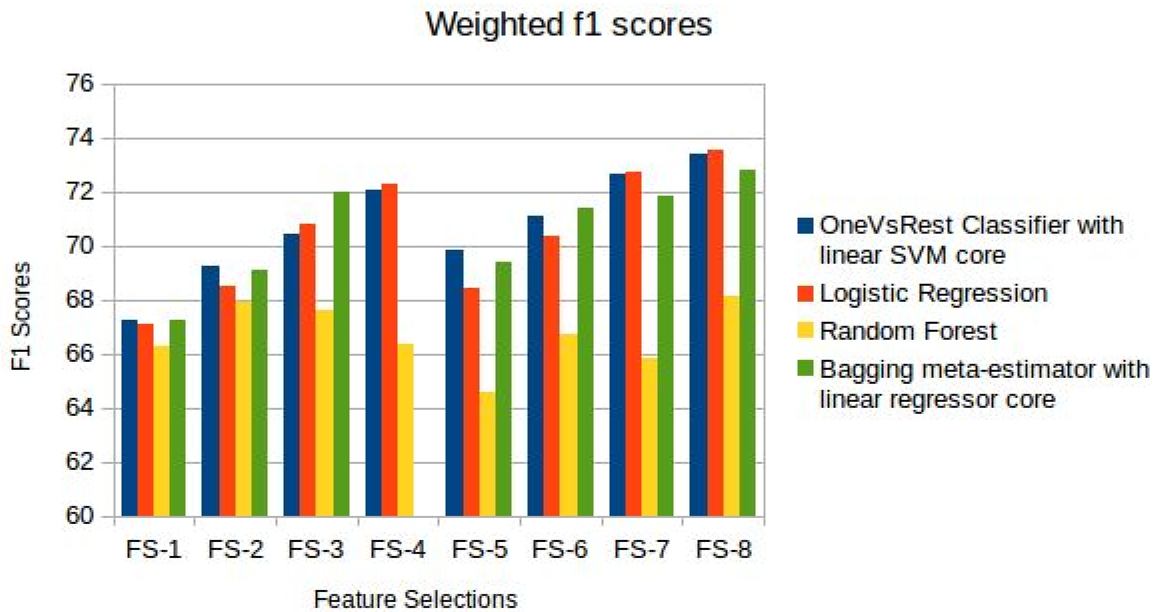


Figure 7.6.2 Various approaches, feature selections and their respective weighted f1-scores

CHAPTER 8:

CONCLUSION

In this concluding chapter, we take a broad look at our system and discuss its merits and demerits. We also expand on how this project can be improved in the future.

8.1 Limitations

- The model trained may not be 100% efficient.
- Some categories of classification may be neglected.
- Automated results may vary from original results.
- Less number of drugs are considered.
- Classification of a medical condition is not provided.
- Automated suggestions for other drugs are not given.

8.2 Conclusion

Thus, we have proposed a novel approach for developing a system for classification and analysis of user reviews for drugs used for treatment of neurological disorders. The approach described above would provide an efficient technique for data collection, annotation and model designing. The proposed system will aim to provide better efficiency as compared to the existing systems. The proposed system would provide an efficient technique for users to enter names of drugs and generate the proposed output. Once implemented, this system would be a boon in the field of medical drug analysis both for the medical professionals as well as the consumers in general.

8.3 Future Scope

The system will be continuously monitored and updated for performance enhancements and bug fixes.
The system could include an interactive real time multi-user forum where users with similar queries can discuss their symptoms. Further, system will provide recommendations for alternative drugs. The system will be extended for other drugs as well with improved efficiency.

9. References

[1] Kuhn, Michael, et al. "A side effect resource to capture phenotypic effects of drugs." Molecular systems biology 6.1 (2010): 343.

[2] Leaman, Robert, et al. "Towards internet-age pharmacovigilance: extracting adverse drug reactions from user posts to health-related social networks." Proceedings of the 2010 workshop on biomedical natural language processing. Association for Computational Linguistics, 2010.

[3] Nikfarjam, Azadeh, et al. "Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features." Journal of the American Medical Informatics Association 22.3 (2015): 671-681.

[4] Yadav, Shweta, et al. "Medical Sentiment Analysis using Social Media: Towards building a Patient Assisted System." LREC. 2018.

[5] Huynh, Trung, et al. "Adverse drug reaction classification with deep neural networks." Coling, 2016.

[6] Yang, Christopher C., et al. "Social media mining for drug safety signal detection." Proceedings of the 2012 international workshop on Smart health and wellbeing. ACM, 2012.

[7] Lee, Kathy, et al. "Adverse drug event detection in tweets with semi-supervised convolutional neural networks." Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2017.

[8] Xia, Long, G. Alan Wang, and Weiguo Fan. "A Deep Learning Based Named Entity Recognition Approach for Adverse Drug Events Identification and Extraction in Health Social Media." International Conference on Smart Health. Springer, Cham, 2017.

[9] Vishal A. Kharde, S.S.Sonawane, “Sentiment Analysis of Twitter Data: A Survey of Techniques” International Journal of Computer Applications, April 2016.

[10] Qurat Tul Ain, Mubashir Ali, Amna Naureen, Muhammad Kamran, “Sentiment Analysis Using Deep Learning Techniques: A Review”, International Journal of Advanced Computer Science and Applications, November 2017.

[11] Vinodhini Gopalakrishnan, Chandrasekaran Ramaswamy, “Patient opinion mining to analyze drugs satisfaction using supervised learning”, Journal of Applied Research and Technology 15 ,2017.

[12] Mark HUGHES, Irene LIa, Spyros KOTOULAS and Toyotaro SUZUMURA, “Medical Text Classification using Convolutional Neural Networks”, International Journal of Computational Linguistics and Applications, April 2016.

13] Aagaard L, Nielsen LH, Hansen EH., “ Consumer reporting of adverse drug reactions: a retrospective analysis of the Danish adverse drug reaction database from 2004 to 2006”, British Journal of Clinical Pharmacology, 2009.

[14] Bian J, Topaloglu[U, Yu F, “Towards large-scale twitter mining for drug-related adverse events”, Proceedings of the 2012 international workshop on smart health and wellbeing, 2012

Project Review Sheet 1

Inhouse/ Industry:

Class: D17 A/B/C

Group No.: 5

Project Evaluation Sheet 2018 - 19

Title of Project: Medical Review Mining using Social Media.

Group Members(sign): Sujay [illegible] Mishra. Dipen Chawla (absent due to medical reasons)

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	4	2	4	2	2	2	2	3	3	2	4	4	42

Comments: Reliability of data is a concern, Need to work on RNN

Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	4	2	3	2	2	2	2	2	2	2	4	4	39

Comments: Authenticity of data need to be considered.

Date: 9th Feb, 2019

Name & Signature Reviewer2

Project Review Sheet 2

Inhouse/ Industry:

Class: D17 A/B/C

Group No.: 05

Project Evaluation Sheet 2018 - 19

Title of Project: Medical Review ~~by~~ using Social Media (Dipen, Sujay, Disha, Varsha)

Group Members(sign): Dipen Sujay [illegible] [illegible]

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	4	2	3	2	2	2	2	2	2	2	4	3	38

Comments: Reliability of data set used is a big conceon

Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg &Mgmt principles (3)	Life - long learning (3)	Professional Skills (5)	Innovative Approach (5)	Total Marks (50)
Review of Project Stage 1	4	4	4	2	3	2	2	1	1	2	2	2	2	[illegible]	[illegible]

Comments:

Date: 6th March, 2019

Name & Signature Reviewer2

10. Appendix

10.1. Paper I

Paper published

A research paper of the system has been submitted on 15th October 2018 for the "*IEEE Technically Co-Sponsored 3rd Biennial International Conference on Nascent Technologies in Engineering"* organized by Fr. C. Rodrigues Institute of Technology, Vashi, Navi Mumbai, India, the conference for which is to be conducted on 04th January 2019, and is pending publishing.

Plagiarism report

PLAGIARISM SCAN REPORT

Words	1968	Date	October 15,2018
Characters	13009	Exclude Url	
7% Plagiarism	93% Unique	6 Plagiarized Sentences	75 Unique Sentences

Content Checked For Plagiarism

Drug review analytics of neurological disorders

Dipen Chawla[1], Disha Mohnani[2], Varsha Sawlani[3], Sujay Varma[4], Sujata Khedkar[5]
Department of Computer Engineering,
V.E.S. Institute of Technology,
Chembur, India
[1]2015dipen.chawla@ves.ac.in, [2]2015disha.mohnani@ves.ac.in, [3]2015varsha.sawlani@ves.ac.in, [4]2015sujay.varma@ves.ac.in,
[5]sujata.khedkar@ves.ac.in

***Abstract*—These days, adverse reactions caused due to medical drugs are one of the major causes of loss of human life. Highly priced laboratory tests aren't enough to obtain all the adverse reactions caused by the majority of the drugs. As a result, it is the need of the hour to develop systems which would supervise effects of drugs after they are cleared for use. Here, we evaluate a self-operating system for drug effectiveness identification on a set of user comments which are annotated manually. We shall try to obtain a relation between the already documented adverse reactions of a drug and those obtained by the proposed system. For this purpose, the system would use unlabeled data. It has been observed that user comments contain a vast variety of complex sentences which pose a natural language challenge. However, these user reviews provide huge scope for further exploration as well.**

Keywords—Medical sentiment, Drug reviews, Pharmacovigilance, Deep learning, Convolutional Neural Network.

I. Introduction

Today's healthcare professionals use online platforms such as blogs, social media, and websites extensively to convey opinions on health matters and the use of drugs. The use of such distributed data available on the World Wide Web, with the ultimate aim of informing public health and policies is termed as 'infodemiology'. On the other side of the healthcare spectrum, users commonly reach out to various social media platforms and blogs in order to convey their views about an administered drug.

Extracting Adverse Drug Reactions (ADE) from such comments provides a new direction to the field of healthcare, which is already in the midst of a data explosion, thanks to the growth of social media as a tool for healthcare analysis.

This paper describes a novel deep learning approach that can be used to analyze patient reviews from social media platforms to generate concise summaries and visualization out of hordes of reviews.

The drug reviews have been analyzed to a certain extent by previous lexicon based approaches. These systems include using a dictionary of words for categorizing user reviews. The proposed system aims to use a deep learning based model rather than the traditional lexicon based approaches, as lexicon based models are incapable of covering the vast subjective and descriptive complexities of user drug reviews. It is quite common for a user to post a review which does not contain generic medical terms, which are at the core of lexicon-based approaches. Rather the reviews contain commonly used non-medical terms which cannot be covered using a traditional dictionary-based approach. A few examples showing the inefficiencies in a lexicon based approach are shown below:

1. "My eyes never adjust to the changes in light and everything becomes a huge swirling mass of shadows." - incorrectly classified as 'EFFECTIVE'

2. "It keeps me from crashing after mania because the mania never peeks." - incorrectly classified as 'OTHER'

3. "It helps with the intense mood swings negative emotions are less severe." - incorrectly classified as 'ADVERSE'

It can be seen from the above examples that review no.1 contains phrases which indicate that the drug has caused severe reactions to the patient's eyes but none of the terms match with those in a dictionary. On the contrary, review numbers 2 and 3 indicate that the drug is effective for the user but they either contain terms related to the adverse effects of the drug or do not contain terms which indicate the effectiveness of the drug. Therefore, they are wrongly classified. These inefficiencies can be overcome by using a non-lexicon based approach.

II. Related Work

Many resources have been manually evaluated for extracting information about post-marketing drug safety, including biomedical literature and electronic health records. However, automatically extracting drug reactions is a complicated task, and very less work has been carried out in said domain. A review of the literature for related work is described below-

Reference [1] introduces a publicly available, computer-friendly resource SIDER that connects 888 drugs to around 1450 side effect terms, as a method for studying Adverse Drug Reactions (ADRs) for academic and commercial research. The publicly available dataset is highly comprehensive in terms of number as well as the quality of reviews and drug effects.

Reference [2] uses a sliding window method to compare terms from a drug reaction lexicon to user comments.

Reference [3] proposes a ML based sequence tagger for automatically extracting ADRs from user reviews. It uses a CRF based classifier which is trained using word embeddings.

Reference [4] introduces a novel approach for sentiment analysis to classify medical conditions and medication using data from social media and training a deep convolutional neural network to obtain highly accurate sentiment word distribution for medical conditions and medication.

Reference [5] uses added attention weights to traditional CNN classifiers, as well as CRNN, a combination of recurrent and convolutional neural networks for adverse reaction extraction.

Reference [6] aims to provide a semi-supervised approach for classifying text and for detecting the adverse reactions of drugs in tweets collected.

In our proposed system, we contrast our deep learning model against standard lexicon based models for efficient classification of user reviews in multiple classes.

The rest of this paper follows this structure: Section III discusses the annotation scheme for user reviews. Section IV discusses the proposed methodology for the extraction of Adverse Reactions. Section V lists the data collection and outcome generation of the model, followed by conclusion in Section VI and references in Section VII.

III. Annotation Scheme

In this section, we define at the first benchmark by studying the views expressed in the user comments. Based on our study, we classify the user reviews into the following four categories:

A. Effective: Indicates that the user has benefitted from a particular drug and so the drug is effective.

B. Non-effective: Indicates that the drug wasn't of any use as the user did not get any benefit from it.

C. Adverse Drug Reaction: Indicates that the user has been affected by the drug and the drug has had an adverse effect on the user.

D. Other: Indicates that the user is expressing any general review that is not related to any medical symptom or medication.

IV. Proposed Model

A. Module 1: Data Scraper/ Web Crawler

This module uses a Scrapy based web crawler for parsing web pages. For this purpose, user comments from three healthcare websites are scraped using a self-designed web crawler. The obtained comments are then stored in the database. We have scraped in all 3000 user reviews from drugs.com, WebMD, everydayhealth.org. Scrapy based web crawler is smart enough to scrape most helpful comments and hence, the crawler will crawl only scrape useful comments.

B. Module 2: Data Preprocessing

The preprocessing module works on the obtained data. Preprocessing includes splitting the individual reviews into sentences, stemming and lemmatization. These steps provide clean data needed for training the model. Descriptive and non-descriptive reviews are separated in this phase.

C. Module 3: Annotation of Data

This phase includes annotating the clean data obtained from the preprocessing stage, only the descriptive reviews are used in this phase. The reviews are annotated into four categories i.e. Effective, Ineffective, Severe adverse reaction and Other.

For this purpose, we make use of an annotation scheme for identifying the drug names and the relationships between the relevant keywords. Keywords indicate the category in which the review under consideration lies. Annotation is done using an annotation tool like Brat. This module provides data which is ready for training

D. Module 4: Classification model

The data is trained using a classification model. The model is based on both deep machine learning approaches and deep learning approaches. Classifiers to be used include Random forest classifier, CRF, CNN, and RNN classifier. The efficiency of each classifier can be compared in order to obtain the most efficient classification.

E. Module 5: Clustering of data

In this module the reviews from the classification phase are taken, clustering of that data is done and the analysis of each segregated category takes place.

F. Module 6: Visualization of data and Report generation

An overall summary report based on the reviews of the drug under consideration is generated in this phase. A report including the pros and cons of the drug is also obtained. A comparison of different drugs for the same disorder can be performed. This could be used for a suggestion of alternative drugs for a drug having severe adverse effects. A report indicating the percentage of satisfied and dissatisfied users can be obtained. Graphical analysis of various drugs with respect to a particular disease can also be obtained.

V. Dataset And Analysis

A. Dataset Preparation

We used drugs.com, everydayhealth.org, webmd.com for extracting user comments. These platforms serve a huge opportunity for patients to share their reviews about drugs which can actually help in discoveries and innovations of existing drugs. Every site has an average of 200 reviews for each drug. We focused on extracting reviews for drugs taken for neurological disorders like seizures, epilepsy, bipolar disorders depending on generic names of drugs defined by USA drug and food administration.

B. Reports, Outcomes, and Analysis

Figure 1.2 shows the expected reports generated by analyzing the percentage of reviews in each of the four categories mentioned previously. This would be helpful in determining the number of satisfied users for the drug under consideration. Figure 1.3 shows the visualization for the comparison of two drugs for the treatment of the same disorder. This would be helpful in suggesting alternative drugs for drugs having severe adverse reactions. The system would also generate a report consisting of a summary of the selected drug. This would be helpful in getting an overall description of the drug based on user comments. Also, the system would generate a report consisting of the pros and cons of the chosen drug.

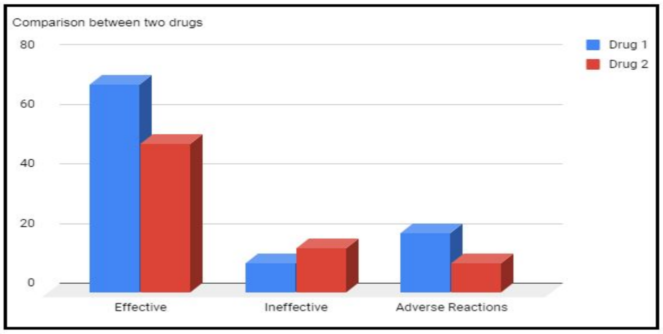


Figure 1.2: Visualization for comparison between two drugs for the same disorder

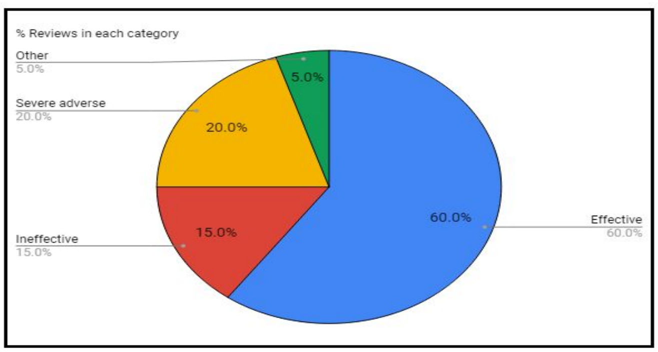


Figure 1.3: Visualization for % reviews in each category

VI. Conclusion

Thus, we have proposed a novel approach for developing a system for classification and analysis of user reviews for drugs used for the treatment of neurological disorders. The proposed system has been designed to provide better efficiency as compared to existing systems. To the best of our knowledge, such a system in this domain has not yet been developed.

References

[1] Kuhn, Michael, Monica Campillos, Ivica Letunic, Lars Juhl Jensen, and Peer Bork. "A side effect resource to capture phenotypic effects of drugs." Molecular systems biology 6, no. 1 (2010): 343.

[2] Leaman, Robert, Laura Wojtulewicz, Ryan Sullivan, Annie Skariah, Jian Yang, and Graciela 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, pp. 117-125. Association for Computational Linguistics, 2010.

[3] Nikfarjam, Azadeh, Abeed Sarker, Karen O'connor, Rachel Ginn, and Graciela Gonzalez. "Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features." Journal of the American Medical Informatics Association 22, no. 3 (2015): 671-681.

[4] Yadav, Shweta, Asif Ekbal, Sriparna Saha, and Pushpak Bhattacharyya. "Medical Sentiment Analysis using Social Media: Towards building a Patient Assisted System." In LREC. 2018.

[5] Huynh, Trung, Yulan He, Alistair Willis, and Stefan Rüger. "Adverse drug reaction classification with deep neural networks." Coling, 2016.

[6] Lee, Kathy, Ashequl Qadir, Sadid A. Hasan, Vivek Datla, Aaditya Prakash, Joey Liu, and Oladimeji Farri. "Adverse drug event detection in tweets with semi-supervised convolutional neural networks." In Proceedings of the 26th International Conference on World Wide Web, pp. 705-714. International World Wide Web Conferences Steering Committee, 2017.

[7] Yang, Christopher C., Haodong Yang, Ling Jiang, and Mi Zhang. "Social media mining for drug safety signal detection." In Proceedings of the 2012 international workshop on Smart health and wellbeing, pp. 33-40. ACM, 2012.

[8] Xia, Long, G. Alan Wang, and Weiguo Fan. "A Deep Learning Based Named Entity Recognition Approach for Adverse Drug Events Identification and Extraction in Health Social Media." In International Conference on Smart Health, pp. 237-248. Springer, Cham, 2017

Participation Certificates

Agnel Charities' Fr. C. Rodrigues Institute of Technology

Vashi, Navi Mumbai 400 703, India

3rd Biennial International Conference on Nascent Technologies in Engineering (ICNTE2019)

Certificate

This is to certify that Dr/Mr/Ms. Dipen Lachhman Chawla

has presented paper titled. Drug review analytics of neurological disorders

in the ICNTE2019 organized by Fr.C. Rodrigues Institute of Technology, Vashi in its premises in association with IEEE on January 4-5, 2019.

IEEE Advancing Technology for Humanity

ICNTE Connecting ideas.

Dr. Nitesh P. Yelve
Conference Chair

Dr. S.M. Khot
Principal

Agnel Charities' Fr. C. Rodrigues Institute of Technology
Vashi, Navi Mumbai 400 703, India

3rd Biennial International Conference on Nascent Technologies in Engineering (ICNTE2019)

Certificate

This is to certify that Dr/Mr/Ms. Disha Mohnani

has presented paper titled Drug review analytics of neurological disorders

in the ICNTE2019 organized by Fr.C. Rodrigues Institute of Technology, Vashi in its premises in association with IEEE on January 4-5, 2019.

IEEE
Advancing Technology for Humanity

ICNTE
Connecting ideas.

Dr. Nitesh P. Yelve
Conference Chair

Dr. S.M. Khot
Principal

Agnel Charities' Fr. C. Rodrigues Institute of Technology
Vashi, Navi Mumbai 400 703, India

3rd Biennial International Conference on Nascent Technologies in Engineering (ICNTE2019)

Certificate

This is to certify that the paper titled..Drug review analytics of neurological disorders... authored by Dr/Mr/Ms ..Varsha Ghanshyam Sawlani... was presented in his/her absentia by his/her co-author during the ICNTE2019 organized by Fr. C. Rodrigues Institute of Technology, Vashi in its premises in association with IEEE on January 4-5, 2019.

IEEE Advancing Technology for Humanity

ICNTE Connecting ideas.

Dr. Nitesh P.
Conference Chair

Dr. S.M. Khot
Principal

10.2. Paper II

Drug review analytics of neurological disorders

Dipen Chawla[1], Disha Mohnani[2], Varsha Sawlani[3], Sujay Varma[4], Sujata Khedkar[5]

Department of Computer Engineering,
V.E.S. Institute of Technology,
Chembur, India

[1]2015dipen.chawla@ves.ac.in, [2]2015disha.mohnani@ves.ac.in, [3]2015varsha.sawlani@ves.ac.in, [4]2015sujay.varma@ves.ac.in,
[5]sujata.khedkar@ves.ac.in

***Abstract*—These days, adverse reactions caused due to medical drugs are one of the major causes of loss of human life. Highly priced laboratory tests aren't enough to obtain all the adverse reactions caused by the majority of the drugs. As a result, it is the need of the hour to develop systems which would supervise the effects of drugs based on user experiences. Here, we develop a system for evaluating drug effectiveness by using a set of user comments which are annotated manually. We have tried to obtain a relation between the already documented adverse reactions of a drug and those obtained by the developed system. For this purpose, the system uses user reviews of those drugs. It has been observed that user comments contain a vast variety of complex sentences which pose a natural language challenge. However, these user reviews provide huge scope for further exploration as well.**

Keywords—Medical sentiment, Drug reviews, Pharmacovigilance, Deep learning, Machine Learning.

I. Introduction

Today's healthcare professionals use online platforms such as blogs, social media, and websites extensively to convey opinions on health matters and the use of drugs. The use of such distributed data available on the World Wide Web, with the ultimate aim of informing public health and policies is termed as 'infodemiology'. On the other side of the healthcare spectrum, users commonly reach out to various social media platforms and blogs in order to convey their views about an administered drug.

Extracting Adverse Drug Reactions (ADR) from such comments provides a new direction to the field of healthcare, which is already in the midst of a data explosion, thanks to the growth of social media as a tool for healthcare analysis.

This paper describes a novel approach that can be used to analyze patient reviews from social media platforms to generate concise summaries and visualization out of hordes of reviews.

The drug reviews have been analyzed to a certain extent by previous lexicon based approaches. These systems include using a dictionary of words for categorizing user reviews. Whereas the system described below makes use of machine learning as well as deep learning models rather than the traditional lexicon based approaches. This is because lexicon based models are incapable of covering the vast subjective and descriptive complexities of user drug reviews. It is quite common for a user to post a review which does not contain generic medical terms, which are the core of lexicon-based approaches. Rather the reviews contain commonly used non-medical terms which cannot be covered using a traditional dictionary-based approach. A few examples showing the inefficiencies in a lexicon based approach are shown below:

1. "My eyes never adjust to the changes in light and everything becomes a huge swirling mass of shadows." - incorrectly classified as 'EFFECTIVE'

2. "It keeps me from crashing after mania because the mania never peeks." - incorrectly classified as 'OTHER'

3. "It helps with the intense mood swings negative emotions are less severe." - incorrectly classified as 'ADVERSE'

It can be seen from the above examples that review no.1 contains phrases which indicate that the drug has caused severe reactions to the patient's eyes but none of the terms match with those in a dictionary. On the contrary, review numbers 2 and 3 indicate that the drug is effective for the user but they either contain terms related to the adverse effects of the drug or do not contain terms which indicate the effectiveness of the drug. Therefore, they are wrongly classified. These inefficiencies can be overcome by using a non-lexicon based approach.

II. Related Work

Many resources have been manually evaluated for extracting information about post-marketing drug safety, including biomedical literature and electronic health records. However, automatically extracting drug reactions is a complicated task, and very less work has been carried out in said domain. A review of the literature for related work is described below-

Reference [1] introduces a publicly available, computer-friendly resource SIDER that connects 888 drugs to around 1450 side effect terms, as a method for studying Adverse Drug Reactions (ADRs) for academic and commercial research. The publicly available dataset is highly comprehensive in terms of number as well as the quality of reviews and drug effects.

Reference [2] uses a sliding window method to compare terms from a drug reaction lexicon to user comments.

Reference [3] proposes a ML based sequence tagger for automatically extracting ADRs from user reviews. It uses a CRF based classifier which is trained using word embeddings.

Reference [4] introduces a novel approach for sentiment analysis to classify medical conditions and medication using data from social media and training a deep convolutional neural network to obtain highly accurate sentiment word distribution for medical conditions and medication.

Reference [5] uses added attention weights to traditional CNN classifiers, as well as CRNN, a combination of recurrent and convolutional neural networks for adverse reaction extraction.

Reference [6] aims to provide a semi-supervised approach for classifying text and for detecting the adverse reactions of drugs in tweets collected.

After researching the extensive work in the automated pharmacovigilance domain, we felt the need for building an accurate system that would be accessible and understandable to the end user for him/her to make an informed decision before consuming the aforementioned drugs.

In our system, we contrast our model against standard lexicon based models for efficient classification of user reviews in multiple classes and provide a simple and interactive interface for conveying the extensive data to the user in a concise manner.

Following the work in this blogpost [9], reviews were extracted, cleaned, preprocessed, analysed, sampled. The inspiration behind creating a OneVsRest classifier in ML and LSTM with multiple variations was obtained in [10].

The rest of this paper follows this structure: Section III discusses the annotation scheme for user reviews. Section IV discusses the methodology followed for developing the system. Section V lists the dataset preparation, the algorithms used and analysis, followed by results obtained from each algorithm using multiple feature sets in Section VI and conclusion and references in Section VII.

III. Annotation Scheme

In this section, we define at the first benchmark by studying the views expressed in the user comments. Based on our study, we classify the user reviews into the following four categories:

A. Effective: Indicates that the user has benefitted from a particular drug and so the drug is effective.

B. Non-effective: Indicates that the drug wasn't of any use as the user did not get any benefit from it.

C. Adverse Drug Reaction: Indicates that the user has been affected by the drug and the drug has had an adverse effect on the user.

D. Other: Indicates that the user is expressing any general review that is not related to any medical symptom or medication.

IV. Model

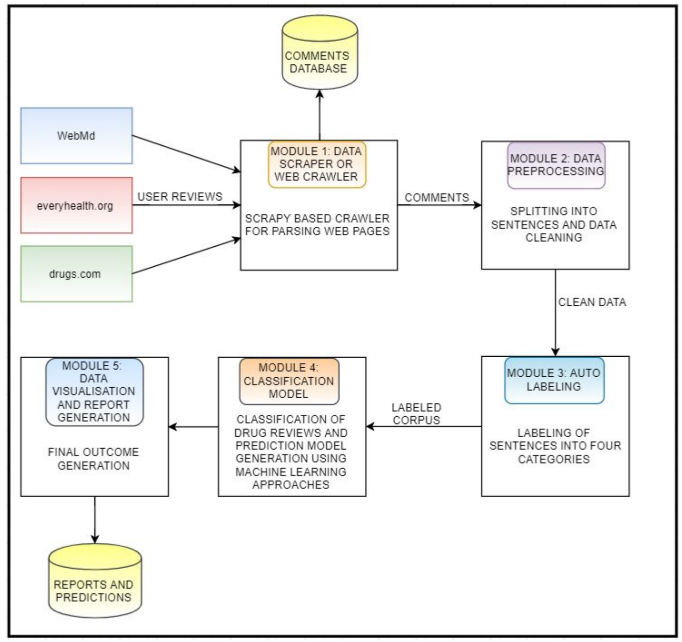


Fig 4.1: Modular diagram of the developed system

A. Module 1: Data Scraper/ Web Crawler

This module uses a Scrapy based web crawler for parsing web pages. For this purpose, user comments from three healthcare websites are scraped using a self-designed web crawler. The obtained comments are then stored in the database. We have scraped in all 3000 user reviews from drugs.com, WebMD, everydayhealth.org.

B. Module 2: Data Preprocessing

The preprocessing module works on the obtained data. Preprocessing includes splitting the individual reviews into sentences, stemming and lemmatization. These steps provide clean data needed for training the model. Descriptive and non-descriptive reviews are separated in this phase.

C. Module 3: Annotation of Data

This phase includes annotating the clean data obtained from the preprocessing stage, only the descriptive reviews are used in this phase. The reviews are annotated into four categories i.e. Effective, Ineffective, Severe adverse reaction and Other. The drugs considered are acetazolamide, carbamazepine, divalproex, ethosuximide, phenobarbital phenytoin, and valporic acid.

An auto-labelling approach was also implemented but was not as effective as the manual classification. As a result all sentences were manually classified to obtain the best results. This module provides data which is ready for training.

D. Module 4: Classification model

The data is trained using a classification model. The model is based on both deep and machine learning and deep learning approaches. Classifiers used include Random forest classifier, SVM classifier, CNN and LSTM approaches. The efficiency of each classifier is compared in order to obtain the most efficient classification. The most accurate prediction is obtained by using a OneVsRest Classifier.

E. Module 6: Visualization of data and Report generation

An overall summary report based on the reviews of the drug under consideration is generated in this phase. A report including the cons of the drug is also obtained. A comparison of different drugs for the same disorder is performed in the form of various charts and graphs. This could be used for obtaining a suggestion of alternative drugs to a drug having severe adverse effects. Besides the top five sentences of each category of every drug are also obtained.

V. Dataset, Algorithms And Analysis

A. Problem Definition

We classify the problem as a multi-class classification problem. The definition of the problem under consideration is given below:

Let R = {r1; r2; ...; rN} be a set of N reviews and C={1; 2; ...; M} be a set of M classes. Given a set of mappings of the from {r_i : c1 ; ...; ck} where review r_i belongs to R, our goal is to find the class for a new review r_{new}.

B. Dataset Preparation

We used drugs.com, everydayhealth.org, webmd.com for extracting user comments. These platforms serve a huge opportunity for patients to share their reviews about drugs which can actually help in discoveries and innovations of existing drugs. Every site has an average of 200 reviews for each drug. We focused on extracting reviews for drugs taken for neurological disorders like seizures, epilepsy, bipolar disorders depending on generic names of drugs defined by USA drug and food administration.

The drugs considered are acetazolamide, carbamazepine, divalproex, ethosuximide, phenobarbital phenytoin, and valporic acid.

C. Feature Selection

The reviews obtained from section A were subjected to a TF-IDF (Term Frequency - Inverse Document Frequency) vectorizer which converts the words in the review to numerical features depending upon how relevant the term is in a given document. We selected the top 15000 terms from our reviews as features for our classification model.

In union with the TF-IDF features, we used sentiment scores generated using VADER library as additional features for the classifier.

Using combinations of these features, we generated 8 feature sets, from FS1 to FS8, for evaluating the model performance thoroughly. These feature sets are further explained in section 6.

C. Model Selection

We compared a total of 6 models for their efficiency using f1 scores for each class, and macro-averaging them, using the support of each class as weights.

We study the usefulness of multi-class classification algorithms like OneVsRest SVM classifier, decision trees, logistic regression, bagging meta-estimator, LSTM, and Convolutional Neural Networks with respect to our problem.

A supervised learning approach is used for classification. Given the feature vectors of training reviews and their corresponding classes as supervision develops a classification model. This model can then be used for predicting the class for a new unclassified review. A summary of how each classifier works is given below:

1. Support Vector Machine

"Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both binary as well as multi-class classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyperplane that differentiates the classes suitably.

2. Random Forest

Random forest classifier uses a multitude of decision trees for classification of data points with the decision for every feature occuring at a single node of a tree with the help of a impurity score. Prediction is made by choosing a class with majority of the voting from all individual decision trees. It can be adapted to multi class environment but is not efficient as well as prone to overfitting.

3. OneVsRest Classifier

Also known as one-vs-all, this strategy consists of fitting one classifier per class. For each classifier, the class is fitted against all the other classes. In addition to its computational efficiency, one advantage of this approach is its interpretability. Since each class is represented by one and one classifier only, it is possible to gain knowledge about the class by inspecting its corresponding classifier. This is the most commonly used strategy for multiclass classification.

4. Bagging Classifier

Bagging Classifier is an ensemble approach which uses it's base classifier (or core) by fitting it on random subsets of data, and then aggregating all their individual predictions. This generally results in lower error rates as the classifier is more generalized over the training set.

D. Testing

Different types of testing performed on the system are:

- *Unit testing*: All the units i.e model development, visualization, summary extraction etc. were tested individually before integration. This ensured that every unit functioned properly without any errors.

- *Integration testing:* Once all the units were developed they were integrated together and tested for any errors. All the individual units developed were integrated using flask framework.

- *Stress testing*: The system was stress tested by varying the number of reviews, from 100 to more than 3000, given for training the model. This would ensure smooth functioning of the system on varied inputs.

The model was found to predict the class of a review very accurately. This showed that the length of the review was not a constraint for prediction. It was also able to learn from thousands of reviews as training data. The system was found to be compatible with both the operating systems considered. It also displayed the exact same user interface on all the browsers considered.

VI. Outcomes and Reports

A. Results

The highest accuracy of the model is obtained by feeding a combination of top 15000 TF-IDF features from the review texts and VADER sentiment scores to a OnevsRestClassifier with a Linear SVM core and cross-validated across 10 folds. This results in an accuracy of ~75% across three categories - adverse, effective, ineffective. The detailed output is shown in table 6.1

Comparison over feature sets -

Eight feature sets from TF-IDF features and VADER sentiment scores were used to evaluate four models as below:

FS-1 : CountVectorizer

FS-2 : CountVectorizer + VADER Sentiment Scores

FS-3 : CountVectorizer top 10000 features + VADER Sentiment Scores + n-gram range 1-3

FS-4 : CountVectorizer all features + VADER Sentiment Scores + n-gram range 1-3

FS-5 : TfidfVectorizer

FS-6 : TfidfVectorizer + VADER Sentiment Scores

FS-7 : Tfidf Vectorizer top 10000 features + VADER Sentiment Scores + n-gram range 1-3

FS-8 : Tfidf Vectorizer top 15000 features + word tokenize analyser + VADER Sentiment Scores + n-gram range 1-3

M1 : OneVsRest Classifier with linear SVM core

M2 : Logistic Regression

M3 : Random Forest Classifier

M4 : Bagging meta-estimator with linear regressor core

The results obtained using these feature sets are

	FS1	FS2	FS3	FS4	FS5	FS6	FS7	FS8
M1	67.21	69.21	70.42	72.07	69.83	71.11	72.61	73.37
M2	67.11	68.48	70.80	72.25	68.43	70.32	72.71	73.56
M3	66.30	67.93	67.58	66.35	64.54	66.70	65.84	68.10
M4	67.21	69.12	71.96	-	69.38	71.37	71.82	72.77

Table 6.1 Various approaches, feature selections and their respective weighted f1-scores

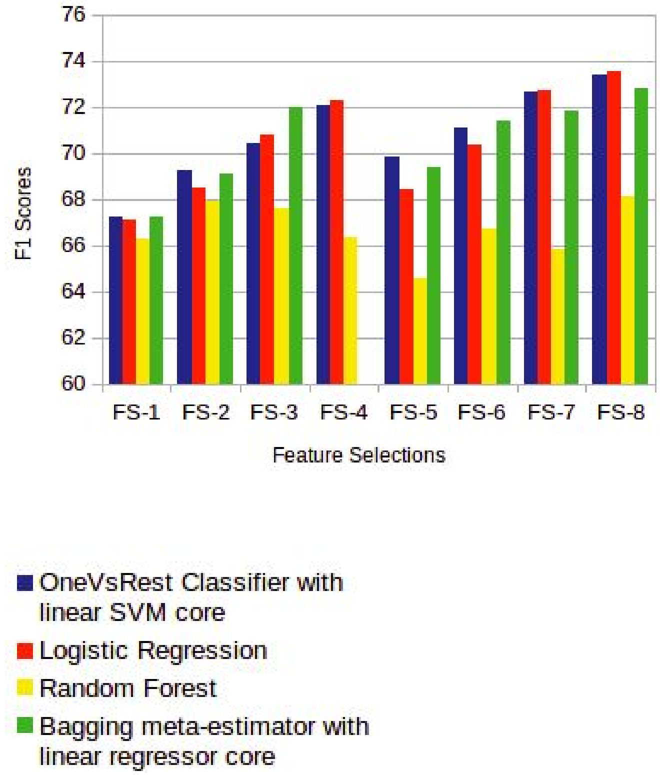


Figure 6.1 Various approaches, feature selections and their respective weighted f1-scores

Various deep learning algorithms like LSTM, CNN + LSTM, LSTM with recurrent dropout were implemented as well. This gave results of around ~75% with high bias towards the dominant class, i.e. effective category. Thus, we conclude that the machine learning approach is more effective as it gives a prediction model which equally favours all classes.

B. Outcomes or Reports

The final web interface for the end user is generated using flask and aids the user to get an concise overview of the drugs he/she intends to use.

The web app allows the user to upload his/her own reactions to the selected drugs, read the reviews in a summarized format or at length, and visualise the drug reactions in an interactive format using graphs and charts.

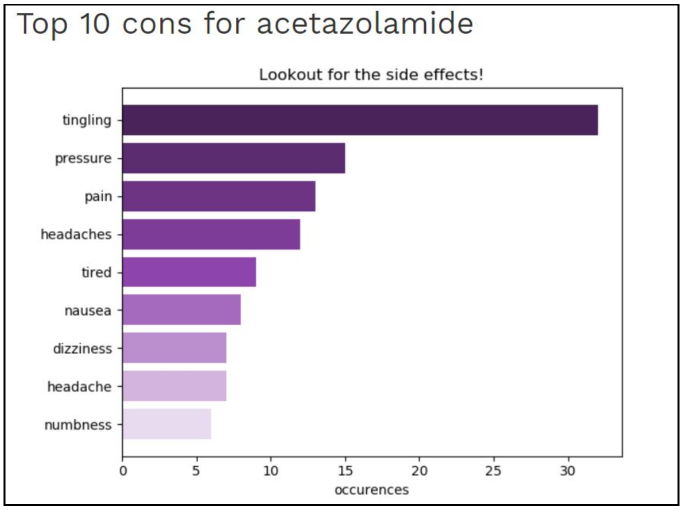


Figure 6.2: Visualization for the adverse reactions for a drug

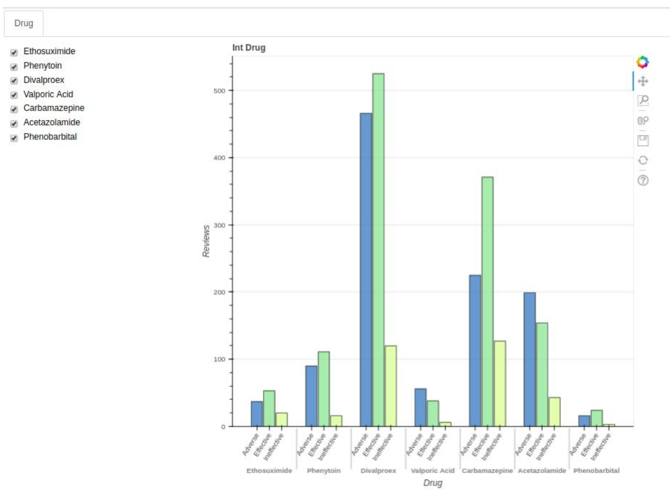


Figure 6.3: Visualization for comparison between drugs for the same disorder

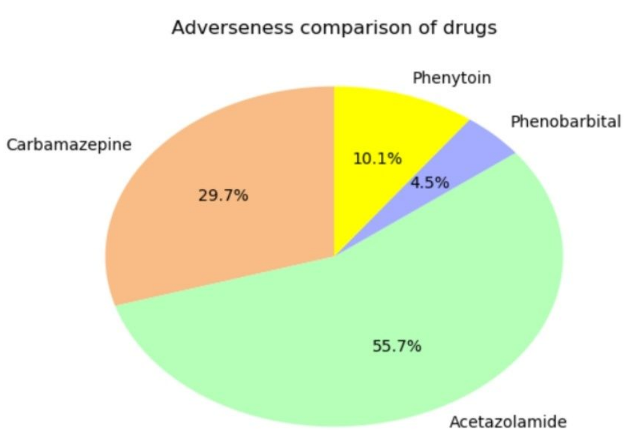


Figure 6.4: Visualization for % adverse reviews in each category

VII. Conclusion

Thus, we have proposed a novel approach for developing a system for classification and analysis of user reviews for drugs used for the treatment of neurological disorders. The proposed system has been designed to provide better efficiency and usability as compared to existing systems. To the best of our knowledge, such a system in this domain has not yet been developed.

References

[1] Kuhn, Michael, Monica Campillos, Ivica Letunic, Lars Juhl Jensen, and Peer Bork. "A side effect resource to capture phenotypic effects of drugs." Molecular systems biology 6, no. 1 (2010): 343.

[2] Leaman, Robert, Laura Wojtulewicz, Ryan Sullivan, Annie Skariah, Jian Yang, and Graciela 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, pp. 117-125. Association for Computational Linguistics, 2010.

[3] Nikfarjam, Azadeh, Abeed Sarker, Karen O'connor, Rachel Ginn, and Graciela Gonzalez. "Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features." Journal of the American Medical Informatics Association 22, no. 3 (2015): 671-681.

[4] Yadav, Shweta, Asif Ekbal, Sriparna Saha, and Pushpak Bhattacharyya. "Medical Sentiment Analysis using Social Media: Towards building a Patient Assisted System." In LREC. 2018.

[5] Huynh, Trung, Yulan He, Alistair Willis, and Stefan Rüger. "Adverse drug reaction classification with deep neural networks." Coling, 2016.

[6] Lee, Kathy, Ashequl Qadir, Sadid A. Hasan, Vivek Datla, Aaditya Prakash, Joey Liu, and Oladimeji Farri. "Adverse drug event detection in tweets with semi-supervised convolutional neural networks." In Proceedings of the 26th International Conference on World Wide Web, pp. 705-714. International World Wide Web Conferences Steering Committee, 2017.

[7] Yang, Christopher C., Haodong Yang, Ling Jiang, and Mi Zhang. "Social media mining for drug safety signal detection." In Proceedings of the 2012 international workshop on Smart health and wellbeing, pp. 33-40. ACM, 2012.

[8] Xia, Long, G. Alan Wang, and Weiguo Fan. "A Deep Learning Based Named Entity Recognition Approach for Adverse Drug Events Identification and Extraction in Health Social Media." In International Conference on Smart Health, pp. 237-248. Springer, Cham, 2017

[9] Using Data Science to help Women make Contraceptive Choices by Krittika Krishnan, Insight Data Science via Medium, October 19, 2018

[10] Movie Review Sentiment Analysis EDA and models by Andrew Lukyanenko via Kaggle.