AI PROJECT

DRUG REVIEW PREDICTION

GUIDE: Deepika ma’am and Nidhi ma’am

TECH WARRIORS

- K. JOSHI MANOHAR (CSE)

- T. MANISH (CSE)

- P. SAMPATH KUMAR (CSE)

- P. JHANSI KUMAR (CSE)

- P. SAI MANIKANTA (CSE)

INTRODUCTION:

* **Python** is an [interpreted](https://en.wikipedia.org/wiki/Interpreted_language), [high-level](https://en.wikipedia.org/wiki/High-level_programming_language), [general-purpose](https://en.wikipedia.org/wiki/General-purpose_programming_language) [programming language](https://en.wikipedia.org/wiki/Programming_language). Created by [Guido van Rossum](https://en.wikipedia.org/wiki/Guido_van_Rossum) and first released in 1991.
* Python's design philosophy emphasizes [code readability](https://en.wikipedia.org/wiki/Code_readability) with its notable use of [significant whitespace](https://en.wikipedia.org/wiki/Off-side_rule).
* Its language constructs and [object-oriented](https://en.wikipedia.org/wiki/Object-oriented_programming) approach aims to help programmers write clear, logical code for small and large-scale projects.
* Python is [dynamically typed](https://en.wikipedia.org/wiki/Dynamic_programming_language) and [garbage-collected](https://en.wikipedia.org/wiki/Garbage_collection_(computer_science)). It supports multiple [programming paradigms](https://en.wikipedia.org/wiki/Programming_paradigm), including [procedural](https://en.wikipedia.org/wiki/Procedural_programming), object-oriented, and [functional programming](https://en.wikipedia.org/wiki/Functional_programming).
* Python interpreters are available for many operating systems. A global community of programmers develops and maintains CPython, an open source reference implementation.
* A non-profit organization, the Python Software Foundation, manages and directs resources for Python and CPython development.
* In computer science, **artificial intelligence (AI),** sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and animals.
* The term "artificial intelligence" is used to describe machines that mimic "cognitive" functions that humans associate with other human minds, such as "learning" and "problem solving".
* Artificial intelligence can be classified into three different types of systems: analytical, human-inspired, and humanized artificial intelligence.
* Analytical AI has only characteristics consistent with [cognitive intelligence](https://en.wikipedia.org/wiki/Cognition), generating a cognitive representation of the world and using learning based on past experience to inform future decisions.
* The traditional problems (or goals) of AI research include [reasoning](https://en.wikipedia.org/wiki/Automated_reasoning), [knowledge representation](https://en.wikipedia.org/wiki/Knowledge_representation), [planning](https://en.wikipedia.org/wiki/Automated_planning_and_scheduling), [learning](https://en.wikipedia.org/wiki/Machine_learning), [natural language processing](https://en.wikipedia.org/wiki/Natural_language_processing), [perception](https://en.wikipedia.org/wiki/Machine_perception) and the ability to move and manipulate objects.
* Many tools are used in AI, including versions of [search and mathematical optimization](https://en.wikipedia.org/wiki/Artificial_intelligence#Search_and_optimization), [artificial neural networks](https://en.wikipedia.org/wiki/Artificial_neural_network), and [methods based on statistics.](https://en.wikipedia.org/wiki/Artificial_intelligence#Probabilistic_methods_for_uncertain_reasoning)
* The AI field draws upon [computer science](https://en.wikipedia.org/wiki/Computer_science), [information engineering](https://en.wikipedia.org/wiki/Information_engineering_(field)), [mathematics](https://en.wikipedia.org/wiki/Mathematics) and many other fields.

OBJECTIVES OF RESEARCH:

* The main objective of this project is to generate a response for the drug by saying whether it is positive, negative or neutral by doing sentimental analysis on the desired columns of the dataset.
* Our study shows that it is feasible to use publicly available natural language processing resources to extract responses from drug id’s.
* The same method can be applied to other sections of the drug label—for example, adverse effects.
* The packages used in this project are nltk and vaderSentiment to get the score of the review based on the Genuinity of the review.
* And the same process will be done on rating too. This help to the customer to opt the drug very easily.Nlp helps to explore automated methods for mapping Adverse Drug Events.

PROBLEM STATEMENT and REVIEW LITERATURE:

* There are lot ups and downs in effectiveness of specific drug in the market at a particular point of time, and this follows a situation where doctors fell in dilemma for suggesting the best medicine for their patients.
* This also effects the sales of those specific drugs which dont show much effectiveness which in turn leads to loss for those Pharmaceutical Companies.
* Who Produces the medicine : Pharma company

Who Suggests the medicine : Doctor

Who consumes the medicine : Patient

Which should be Effective in Maintaining a Healthy chain Among these 3 : Medicine/Drug.

* If there is an imbalance in this cycle among company,Doctor and Patient, it leads to loss of money and time for patient ,loss of Reputation of Doctor and Hospital ,and loss in sales for Pharma Company.
* So, we can reduce this imbalance by analysing the drug trends time to time and give an analysis report and overall feedback about the performance of the drug in a particular week to Both Doctor and Pharma companies.

* The main point is which is the main source data for this analysis on Drug - undoubtedly its Patients.
* Tell The Drug is a trend analyser , which runs on Sentimental analysis segment of Artificial intelligence
* How it works ?
* There will be the drug feedback machine’s present at pharmacies across a particular city and each pharmacy in the city is assigned with a particular ID. if a consumer purchases a particular drug at a pharmacy, then he will be able to give feedback about a particular medicine which he purchases.

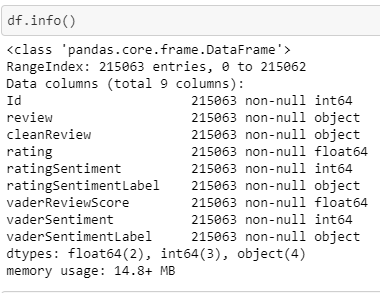
To extract drug indications from structured

drug labels and represent the information using codes

from standard medical terminologies.

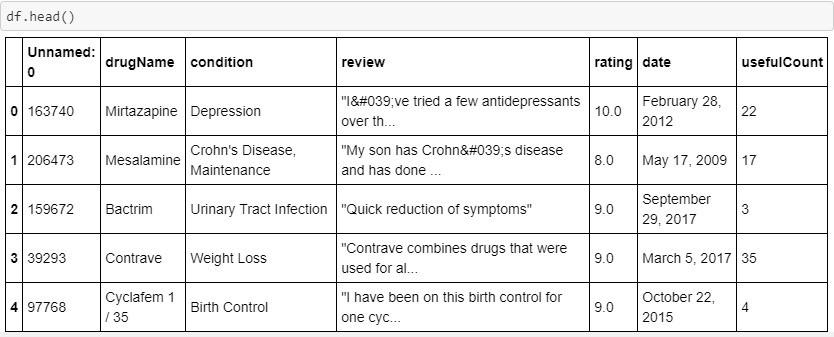
DATA COLLECTION:

* Initially we have two datasets i.e. drugcom\_test & drugcom\_train. we have concatenated the two datasets into a single dataset to get a required output.
* Now the dataset contain 7 columns such as drugid, drugname, review, rating, date, counting labels, sentiment labels.

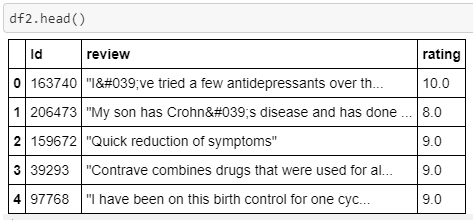


METHODOLOGY and DATA MODELLING:

* This Data set is not the ‘dirtiest’, It’s tidy in ‘.tsv’ files,divided into train and test set and has no Null Values but to build a working Model, Further processing and wrangling has to be done.
* The First step is to load data: There are many ways to load Data into a Notebook.
* *It could be directly through the File upload Icon,*
* *Using ‘!wget url’,*
* For this Exercise, the ‘ !wget ’ command method was used, ‘!unzip ’ to decompress the zipped file, then the files were read into Pandas dataFrames.
* I combined the two files(Test and Train) that were unzipped with “pd.concat“.This is to have a larger data set (We will split them later, maybe in another proportion) and to preprocess the files together.

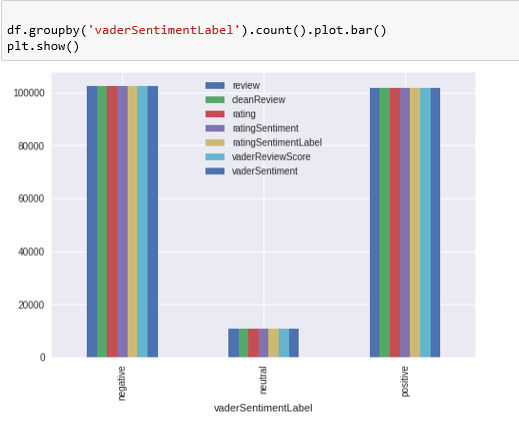


* After inspecting the dataFrames, the columns were renamed.
* A new dataFrame was created containing just the Id,review and rating columns.

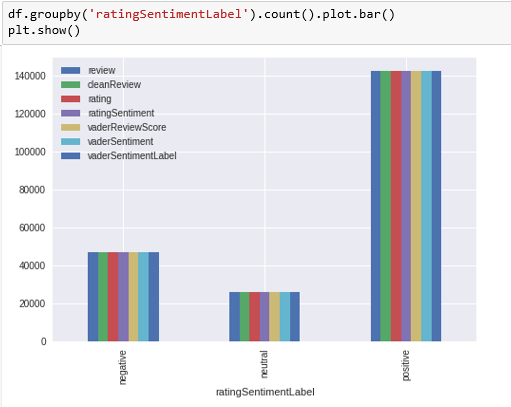


* vaderSentiment library was pip installed
* Stopwords were downloaded from NLTK. You can create your own list of stopwords or use another precompiled list.
* vaderSentiment’s Sentiment Intensity Analyzer was used to generate the compound sentiment polarity scores based on the review.
* The sentiment polarity scores were grouped into 3 polarity labels(Many analysis are based on binary (positive or negative) but I chose the(positive,neutral and negative) labels because in practice, people can be neutral about their medications and I will like that to be learned by the model .
* Readers can choose the binary polarity label, the vaderReviewScore can be regrouped to just two Labels and mapped as a positive or negative or neutral.
* The punctuation in the review were not removed because VADER was designed to be able to analyze Social media data and factors like exclamation marks and special characters are considered in calculating the polarity scores.
* Sentiment Score of 2 for positive, 1 for negative, 0 for neutral were mapped on the vaderReviewScore column as vaderSentiment.
* The vaderSentiment were labelled positive,negative or negative as vaderSentimentLabel
* A similar analysis was carried out on the rating. Sentiment generated from review is considered implicit while the rating is considered explicit. It is not unusual for their polarity labels to be different.
* For eg someone may review a particular item in a positive tone but still give it an average rating while some can review an item in a seemingly negative tone and rate the product highly.
* This is why it is important to infer sentiment from review(using Libraries like VADER) even though it’s easier and straight forward to analyze the star rating values.
* In this project we used vaderSentiment analysis and NLP (natural learning processing) for its better implementation.

FIGURES and GRAPHS:



Vader Sentiment Label



Rating Sentiment Label

CONCLUSIONS:

* Faster patient access to new drugs with superior efficacy for certain subgroups of patients.
* There are lot ups and downs in effectiveness of specific drug in the market at a particular point of time,
* and this follows a situation where doctors fell in dilemma for suggesting the best medicine for their patients.
* This also effects the sales of those specific drugs which dont show much effectiveness which in turn leads to
* loss for those Pharmaceutical Companies.
* Who Produces the medicine : Pharma company
* Who Suggests the medicine : Doctor
* Who consumes the medicine : Patient
* Which should be Effective in Maintaining a Healthy chain Among these 3 : Medicine/Drug.
* So, we can reduce this imbalance by analysing the drug trends time to time and give an analysis report and overall
* feedback about the performance of the drug in a particular week to Both Doctor and Pharma companies.
* main point is which is the main source data for this analysis on Drug.
* undoubtedly its Patients.
* Tell The Drug is a trend analyser , which runs on Sentimental analysis segment of Artificial intelligence
* How it works ?
* There will be Tell the drug feedback machienes present at pharmacies across a particular city and each pharmacy
* in the city is assigned with a particular ID. if a consumer purchases a particular drug at a pharmacy, then he will be able to give
* feedback about a particular medicine which he purchases.

An important objective To extract drug indications from structured

drug labels and represent the information using codes

from standard medical terminologies.

To extract drug indications from structured

drug labels and represent the information using codes

from standard medical terminologies.

To extract drug indications from structured

drug labels and represent the information using codes

from standard medical terminologies.

To extract drug indications from structured

drug labels and represent the information using codes

from standard medical terminologies.

To extract drug indications from structured

drug labels and represent the information using codes

from standard medical terminologies.