JOURNAL OF CRITICAL REVIEWS

ISSN- 2394-5125 VOL 06, ISSUE 07, 2019

Drug Recommendation System based on Sentiment Analysis of Drug Reviews using Multi-Layer Perception

Sharath Pokala, Sathish Parvatham, Bandari Nithya

Department of Electronics and Communication Engineering

Sree Dattha Group of Institutions, Hyderabad, Telangana, India.

Abstract

Since coronavirus has shown up, inaccessibility of legitimate clinical resources is at its peak, like the shortage of specialists and healthcare workers, lack of proper equipment and medicines etc. The entire medical fraternity is in distress, which results in numerous individual's demise. Due to unavailability, individuals started taking medication independently without appropriate consultation, making the health condition worse than usual. As of late, machine learning has been valuable in numerous applications, and there is an increase in innovative work for automation. This work intends to present a drug recommender system that can drastically reduce specialist's heap. To overcome from above problem author of this paper introducing sentiment and machine learning based drug recommendation system which will accept disease names from patient and then recommend DRUG and simultaneously display SENTIMENT rating based on reviews given by old users based on their experience. This work introduces the multi-layer perception (MLP) based drug recommendation system. The term frequency – inverse document frequency (TF-IDF) feature extraction method is used to deep features from the pre-processed dataset. The MLP classifier with TF-IDF feature extraction will result in superior performance compared to other models. To implement this work, DRUGREVIEW dataset was used from UCI machine learning website. Finally, the simulations revealed that the proposed TF-IDF and MLP resulted in superior performance as compared to UCI model.

Keywords: Drug Recommendation System, Multi-Layer Perception. term frequency – inverse document frequency.

1. Introduction

One of the most concerned and searched topics on the internet is about health information. According to the Pew Internet and American Life Project, almost 60% of grownups are looking for enough health information on the web with 35% of respondents concentrating on diagnosing ailments online only. Since many studies show that number of people die due to the medical errors and the semi errors are caused by medical practitioners, who prescribe medicines based on their experiences. As most of their experiences are limited, they often commit mistakes. This study provides a medicine recommendation system for doctors which can be used by them while prescribing medicines. A recommender framework is an ordinary framework that makes the users get a proposal of things which they can utilize for their exact need. Dissimilar to numerous different kinds of frameworks, health recommendation principally relies upon enthusiastic, physical and mental issues of the patients. A medicine recommendation system is similar system that recommend the medicines for a particular disease based on patient reviews. This system is very essential in this fast-growing technological world, which can save lives by helping doctors. In this paper, the proposed medicine recommendation system and its working is depicted, wherein it uses the current technologies like machine learning, data mining etc. to find out the interesting records hidden in the medical data and reduce the medical errors by the doctors while prescribing medicines. This system consists of following modules such as database module, data preparation, data visualization, recommendation, and model evaluation module. The proposed medication recommender system uses Machine learning N-Gram and Lightgbm

algorithms by using data from hospital and the best one is selected for the medicine recommendation system to attain the metrics like good accuracy, scalability, and model efficiency.

Online consultations require the patient to describe their symptoms to the doctor. A spike in virtual medical services has been reported in the wake of the novel coronavirus disease (COVID-19) [1]. Diabetes, hypertension, and heart disease are all associated with an increased risk of virus infections. The availability of health care professionals 24/7, no need for travel, security, privacy, and drug recommendations are all advantages of virtual medical services. The recommender system allows for improvements in medical services in disparate areas [2]. Often, finding a physician in remote areas can be tricky, so recommender systems have been created to help. Health-related recommender systems can make an early diagnosis, predict disease progression, and make appropriate recommendations according to the health status of patients [3,4]. Machine learning (ML) greatly improves the quality of medical recommender systems by providing suggestions that are based on patient needs and feedback [5,6]. By using sentiment analysis and feature engineering, the drug recommender system can dispense medicine according to a specific condition. Emotions, such as attitudes and opinions, are separated and extracted from language through sentiment analysis [7]. By using the recommender system, information overload can be solved, and egovernment and e-learning can be improved [8]. Depending on an individual's health status, these recommender systems prescribe medications, diagnose diseases, and refer them to the relevant health care. An ML-driven recommendation system generates appropriate recommendations using parameters such as blood pressure, gender, cholesterol levels, and blood sugar for diseases such as colds, fevers, and cardiac deaths [9].

2. Literature survey

Bartlett et. al [1] compares on real data effective duplicates detection methods for automatic deduplication of files based on names, working with French texts or English texts, and the names of people or places, in Africa or in the West. After conducting a more complete classification of semantic duplicates than the usual classifications, they introduce several methods for detecting duplicates whose average complexity observed is less than O(2n). Through a simple model, they highlight a global efficacy rate, combining precision and recall. We propose a new metric distance between records, as well as rules for automatic duplicate detection. Analyses made on a database containing real data for an administration in Central Africa, and on a known standard database containing names of restaurants in the USA, have shown better results than those of known methods, with a lesser complexity. Shimada et. al [2] developed a decision support system that helps doctors select appropriate first-line drugs. The system classifies patients' abilities to protect themselves from infectious diseases as a risk level for infection. In an evaluation of the prototype system, the risk level it determined correlated with the decisions of specialists. The system is very effective and convenient for doctors to use.

He et. al [3] presented a novel adaptive synthetic (ADASYN) sampling approach for learning from imbalanced data sets. The essential idea of ADASYN is to use a weighted distribution for different minority class examples according to their level of difficulty in learning, where more synthetic data is generated for minority class examples that are harder to learn compared to those minority examples that are easier to learn.

Lei et. al [4] presented a novel approach to polarity classification of short text snippets, which takes into account the way data are naturally distributed into several topics in order to obtain better classification models for polarity. This approach is multi-step, where in the initial step a standard topic classifier is learned from the data and the topic labels, and in the ensuing step several polarity

ISSN- 2394-5125 VOL 06, ISSUE 07, 2019

classifiers, one per topic, are learned from the data and the polarity labels. They empirically show that our approach improves classification accuracy over a real-world dataset by over 10%, when compared against a standard single-step approach using the same feature sets. The approach is applicable whenever training material is available for building both topic and polarity learning models. Nikfarjam and Gonzalez et. al [5] presented a new method for using association rules for colloquial text mining. They applied our method on user comments to find mentions of adverse reactions to drugs by extracting frequent patterns. Since we are dealing with highly informal colloquial text, the idea of using extracted patterns might, at first, seem counter intuitive. However, we indeed found consistencies in the user comments. This evaluation measured the effectiveness of this technique in extracting frequent patterns in this context. However, this method can easily be generalized for other contexts and languages.

Doulaverakis et. al [6] presented a drug recommendation system based on Semantic Web technologies, termed GalenOW. It has been shown that OWL and Semantic Web technologies can provide a good match for drug recommendations as OWL is expressive enough to effectively encapsulate medical knowledge. Rule-based reasoning can model medical decision making and aid experts. A comparison of the semantic-enabled implementation to a traditional business logic implementation was presented. Although the latter has shown better performance in time and memory requirements, semantic technologies provide a better alternative for integrating knowledge in the system than simple rule engines.

Goeuriot et. al [7] presented creation of lexical resources and their adaptation to the medical domain. We first describe the creation of a general lexicon, containing opinion words from the general domain and their polarity. Then they presented the creation of a medical opinion lexicon, based on a corpus of drug reviews. They show that some words have a different polarity in the general domain and in the medical one. Some words considered generally as neutral are opinionated in medical texts. They finally evaluate the lexicons and show with a simple algorithm that using our general lexicon gives better results than other well-known ones on our corpus and that adding the domain lexicon improves them as well.

Keers et. al [8] appraised empirical evidence relating to the causes of medication administration errors (MAEs) in hospital settings. Limited evidence from studies included in this systematic review suggests that MAEs are influenced by multiple systems factors, but if and how these arise and interconnect to lead to errors remains to be fully determined. Further theoretical focused is needed to investigate the MAE causation pathway, with an emphasis on ensuring interventions designed to minimise MAEs target recognised underlying causes of errors to maximise their impact.

Wittich et. al [9] provides a practicing physician that focuses on medication error terminology and definitions, incidence, risk factors, avoidance strategies, and disclosure and legal consequences. A medication error is any error that occurs at any point in the medication use process. It has been estimated by the Institute of Medicine that medication errors cause 1 of 131 outpatient and 1 of 854 inpatient deaths. Medication factors (eg, similar sounding names, low therapeutic index), patient factors (eg, poor renal or hepatic function, impaired cognition, polypharmacy), and health care professional factors (eg, use of abbreviations in prescriptions and other communications, cognitive biases) can precipitate medication errors.

Zhang et. al [10] proposed a novel cloud-assisted drug recommendation (CADRE), which can recommend users with top-N related medicines according to symptoms. In CADRE, they first cluster the drugs into several groups according to the functional description information and design a basic personalized drug recommendation based on user collaborative filtering. Then, considering the

ISSN- 2394-5125 VOL 06, ISSUE 07, 2019

shortcomings of collaborative filtering algorithm, such as computing expensive, cold start, and data sparsity, they propose a cloud-assisted approach for enriching end-user Quality of Experience (QoE) of drug recommendation, by modeling and representing the relationship of the user, symptom, and medicine via tensor decomposition. Finally, the proposed approach is evaluated with experimental study based on a real dataset crawled from Internet.

Danushka et. al [11] proposed an unsupervised method for learning domain-specific word representations that accurately capture the domain-specific aspects of word semantics. First, we select a subset of frequent words that occur in both domains as \emph{pivots}. Next, they optimize an objective function that enforces two constraints: for both source and target domain documents, pivots that appear in a document must accurately predict the co-occurring non-pivots, and, word representations learnt for pivots must be similar in the two domains. Moreover, they propose a method to perform domain adaptation using the learnt word representations. This proposed method significantly outperforms competitive baselines including the state-of-the-art domain-insensitive word representations and reports best sentiment classification accuracies for all domain-pairs in a benchmark dataset.

3. Proposed system

A recommender framework is a customary system that proposes an item to the user, dependent on their advantage and necessity. These frameworks employ the customers' surveys to break down their sentiment and suggest a recommendation for their exact need. In the drug recommender system, medicine is offered on a specific condition dependent on patient reviews using sentiment analysis and feature engineering. Sentiment analysis is a progression of strategies, methods, and tools for distinguishing and extracting emotional data, such as opinion and attitudes. On the other hand, featuring engineering is the process of making more features from the existing ones; it improves the performance of models.

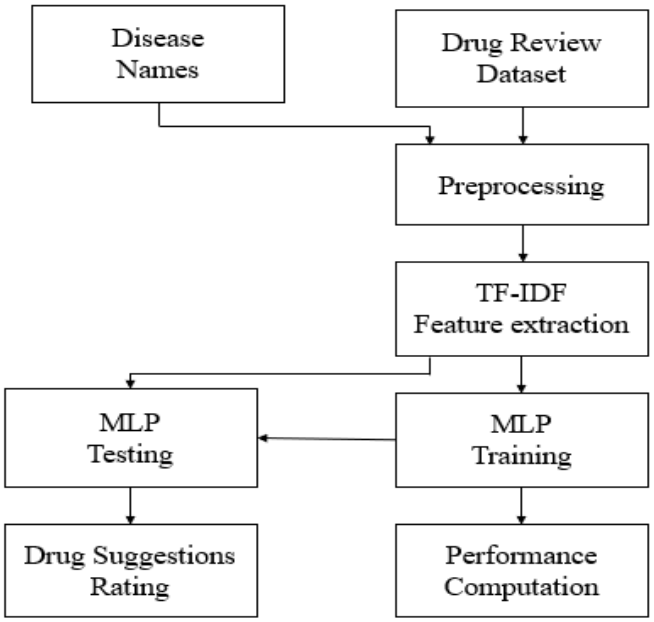

Figure 1. Block Diagram of Proposed System.

The proposed work has used term frequency – inverse document frequency (TF-IDF) feature extraction methods as shown in Figure 1. The MLP classifier with TF-IDF feature extraction will result in superior performance compared to other models. To implement this work, DRUGREVIEW

dataset was used from UCI machine learning website. Finally, the simulations revealed that the proposed TF-IDF and MLP resulted in superior performance as compared to UCI model.

3.1 DRUGREVIEW Dataset

The dataset provides patient reviews on specific drugs along with related conditions and a 10-star patient rating reflecting overall patient satisfaction. The data was obtained by crawling online pharmaceutical review sites. The intention was to study

1. sentiment analysis of drug experience over multiple facets, i.e. sentiments learned on specific aspects such as effectiveness and side effects,
2. the transferability of models among domains, i.e. conditions, and
3. the transferability of models among different data sources (see 'Drug Review Dataset (Druglib.com)').

3.2 Preprocessing

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data preprocessing task.

Need of Data Preprocessing: A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data preprocessing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

3.3 Splitting the Dataset

In machine learning data preprocessing, we divide our dataset into a training set and test set. This is one of the crucial steps of data preprocessing as by doing this, we can enhance the performance of our machine learning model. Suppose if we have given training to our machine learning model by a dataset and we test it by a completely different dataset. Then, it will create difficulties for our model to understand the correlations between the models. If we train our model very well and its training accuracy is also very high, but we provide a new dataset to it, then it will decrease the performance. So we always try to make a machine learning model which performs well with the training set and also with the test dataset.

3.4 TF-IDF Feature extraction

TF-IDF which stands for Term Frequency – Inverse Document Frequency. It is one of the most important techniques used for information retrieval to represent how important a specific word or phrase is to a given document. Let’s take an example, we have a string or Bag of Words (BOW) and we have to extract information from it, then we can use this approach.

The tf-idf value increases in proportion to the number of times a word appears in the document but is often offset by the frequency of the word in the corpus, which helps to adjust with respect to the fact that some words appear more frequently in general. TF-IDF use two statistical methods, first is Term Frequency and the other is Inverse Document Frequency. Term frequency refers to the total number of times a given term t appears in the document doc against (per) the total number of all words in the document and the inverse document frequency measure of how much information the word provides.

ISSN- 2394-5125 VOL 06, ISSUE 07, 2019

It measures the weight of a given word in the entire document. IDF show how common or rare a given word is across all documents.

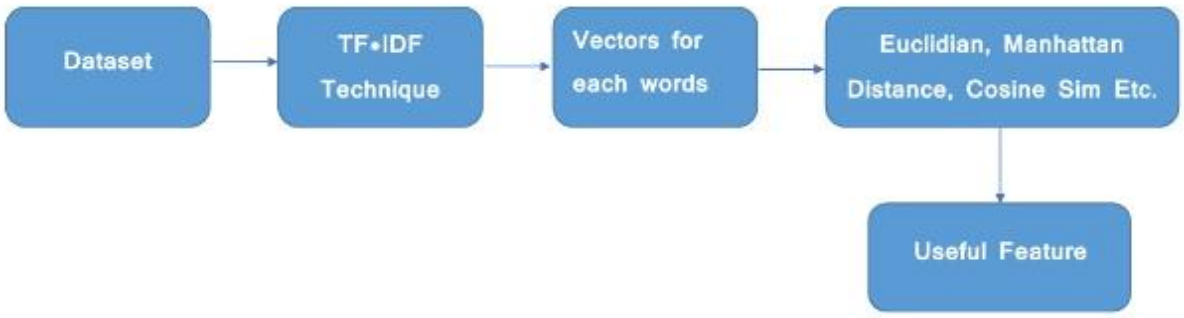

Figure 2. TF-IDF block diagram.

TF-IDF do not convert directly raw data into useful features. Firstly, it converts raw strings or dataset into vectors and each word has its own vector. Then we'll use a particular technique for retrieving the feature like Cosine Similarity which works on vectors, etc.

Term Frequency (TF): Suppose we have a set of English text documents and wish to rank which document is most relevant to the query, "Data Science is awesome!" A simple way to start out is by eliminating documents that do not contain all three words "Data" is", "Science", and "awesome", but this still leaves many documents. To further distinguish them, we might count the number of times each term occurs in each document; the number of times a term occurs in a document is called its term frequency. The weight of a term that occurs in a document is simply proportional to the term frequency.

$$tf(t,d) = count\ of\ t\ in\ d\ /\ number\ of\ words\ in\ d$$

Document Frequency: This measures the importance of document in whole set of corpora, this is very similar to TF. The only difference is that TF is frequency counter for a term t in document d, whereas DF is the count of occurrences of term t in the document set N. In other words, DF is the number of documents in which the word is present. We consider one occurrence if the term consists in the document at least once, we do not need to know the number of times the term is present.

$$df(t) = occurrence\ of\ t\ in\ documents$$

Inverse Document Frequency (IDF): While computing TF, all terms are considered equally important. However, it is known that certain terms, such as "is", "of", and "that", may appear a lot of times but have little importance. Thus, we need to weigh down the frequent terms while scale up the rare ones, by computing IDF, an inverse document frequency factor is incorporated which diminishes the weight of terms that occur very frequently in the document set and increases the weight of terms that occur rarely. **The** IDF is the inverse of the document frequency which measures the informativeness of term t. When we calculate IDF, it will be very low for the most occurring words such as stop words (because stop words such as "is" is present in almost all of the documents, and N/df will give a very low value to that word). This finally gives what we want, a relative weightage.

$$idf(t) = N/df$$

Now there are few other problems with the IDF, in case of a large corpus,say 100,000,000 , the IDF value explodes , to avoid the effect we take the log of idf . During the query time, when a word which

ISSN- 2394-5125 VOL 06, ISSUE 07, 2019

is not in vocab occurs, the df will be 0. As we cannot divide by 0, we smoothen the value by adding 1 to the denominator.

$$idf(t) = log(N/(df + 1))$$

The TF-IDF now is at the right measure to evaluate how important a word is to a document in a collection or corpus. Here are many different variations of TF-IDF but for now let us concentrate on this basic version.

$$tf - idf(t,d) = tf(t,d) * log(N/(df + 1))$$

Implementing TF-IDF: To make TF-IDF from scratch in python, let's imagine those two sentences from different document:

first_sentence: "Data Science is the sexiest job of the 21st century".

second_sentence: "machine learning is the key for data science".

First step we have to create the TF function to calculate total word frequency for all documents.

3.5 Multilayer perceptron

MLP is one of the most frequently used neural network architectures in MDSS and it belongs to the class of supervised neural networks. The multilayer perceptron consists of a network of nodes (processing elements) arranged in layers. A typical MLP network consists of three or more layers of processing nodes: an input layer that receives external inputs, one or more hidden layers, and an output layer which produces the classification results (Fig. 3). Note that unlike other layers, no computation is involved in the input layer. The principle of the network is that when data are presented at the input layer, the network nodes perform calculations in the successive layers until an output value is obtained at each of the output nodes. This output signal should be able to indicate the appropriate class for the input data. That is, one can expect to have a high output value on the correct class node and low output values on all the rest. A node in MLP can be modeled as an artificial neuron, which computes the weighted sum of the inputs at the presence of the bias and passes this sum through the activation function. The whole process is defined as follows

$$v_j = \sum_{i=1}^{p} w_{ji} x_i + \theta_j$$
$$y_j = f_j(v_j)$$

where vj is the linear combination of inputs x1; x2; xp, qj is the bias, wji is the connection weight between the input xi and the neuron j, and fj ($) is the activation function of the jth neuron, and yj is the output.

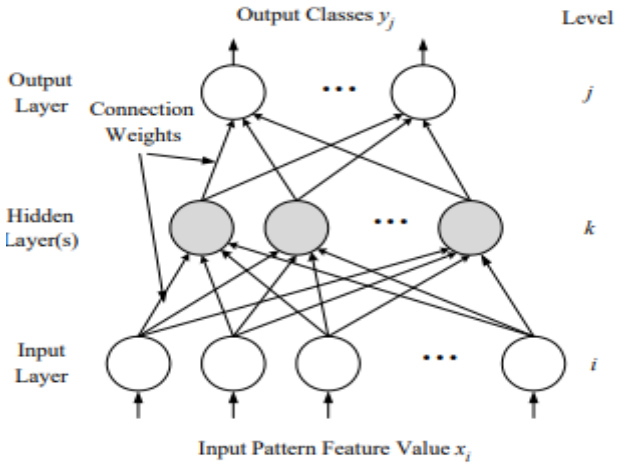

Figure 3. Architecture of a multilayer perceptron network.

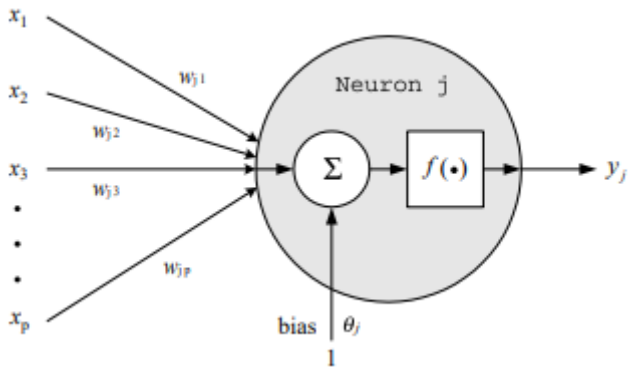

Figure 4. One node of MLP: an artificial neuron.

The sigmoid function is a common choice of the activation function, as defined

$$f(a) = \frac{1}{1 + e^{-a}}$$

The bias term qj contributes to the left or right shift of the sigmoid activation function, depending on whether qj takes a positive or negative value. Once the architecture of MLP has been determined, the connection weights of the network have to be computed through a training procedure based on the training patterns and the desired output. BP is one of the simplest and most general methods for the supervised training of MLP. The error E can be chosen as the mean square error (MSE) function between the actual output yj and the desired output dj:

$$E = \frac{1}{2} \sum_{j=1}^{n_j} (d_j - y_j)^2$$

There are two common training strategies: the incremental training strategy and the batch training strategy. Usually, an incremental strategy is more efficient and also faster for systems with large training samples, as random disturbances can be induced to help the system to escape from a local minimum point. The BP algorithm described above has some shortcomings. If the learning rate is set small enough to minimize the total error, the learning process will be slowed down. On the other hand, a larger learning rate may speed up learning process at the risk of potential oscillation. Another problem is that, partial minimal points or stable stages on error surface are often encountered during

ISSN- 2394-5125 VOL 06, ISSUE 07, 2019

the learning process (Baba, 1989). Using a momentum term is the simplest method to avoid oscillation problems during the search for the minimum value on the error surface. The weight update in BP algorithm with a momentum term a is defined as follows:

$$\Delta W(t) = -\eta \frac{\partial E(t)}{\partial W} + \alpha \Delta W(t-1)$$

where 0! a!1. The adaptive learning rate can also be adopted to speed up the convergence of the algorithm. For batch training strategy, the learning rate can be adjusted as follows

$$\eta(t) = \begin{cases} \beta\eta(t-1) & \text{if } E(t) < E(t-1) \\ \theta\eta(t-1) & \text{if } E(t) > kE(t-1) \\ \eta(t-1) & \text{otherwise} \end{cases}$$

where h(t) is the learning rate at the tth iteration, and b, q, and k are chosen as such that bO1, 0! q!1, and kO1. While for the incremental training strategy, learning rate can be updated using

$$\eta(t) = \eta_0 + \lambda E(t-1)$$

where h0 is a present learning rate. The learning algorithm with forgetting mechanics is an algorithm that can 'forget' unused connections (Takeshi, 2001). With this forgetting mechanism, the weights that are not reinforced by learning will disappear. The obtained network, thus, has a skeletal structure that reflects the regularity contained in the data, useful to improve the convergence and the network accuracy. In general, the updating of connection weights with forgetting mechanics term is given by:

$$\Delta W'(t) = \Delta W(t) - \varepsilon \operatorname{sgn}(W(t))$$

where 3 is the amount for the forgetting, and sgn(x) is the sign function (i.e., sgn(x)Z1 if xO0, K1 if x!0, and 0 if xZ0). The absolute value of connection weight is set to decrease by 3 due to the second term on the right-hand side. In practice, some optimization algorithms are often used to improve the network convergence, such as the steepest descent method, the Newton method, the Quasi–Newton method, and the conjugate gradients method. In this study, the conjugate gradients method is adopted, as it has a low computation cost and exhibits good results. The connection weights thus can be expressed by:

$$W(t+1) = W(t) + \eta(t)d(t)$$

$$d(t) = -\nabla E[W(t)] + \beta(t)d(t-1)$$

$$d(0) = -\nabla E[W(0)]$$

where PE is the gradient, d(t) is conjugate gradient, h(t) is the step wide, b(t) is determined given by Polak–Ribiere function

$$\beta(t) = \frac{[\nabla E(W(t)) - \nabla E(W(t-1))]^T \nabla E[W(t)]}{\nabla E[W(t-1)]^T \nabla E[W(t-1)]}$$

4. Results

This section gives the detailed analysis of simulation results implemented using "python environment". Further, the performance of proposed method is compared with existing methods using same dataset.

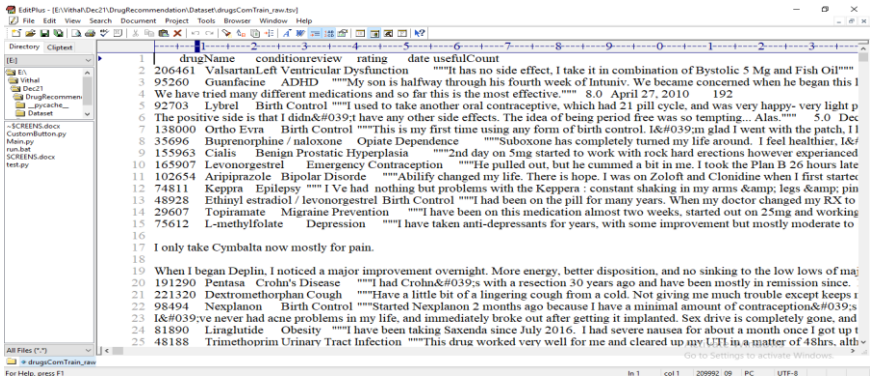

Figure 5. Sample dataset.

In Figure 5, first row represents dataset column names such as drug name, condition, review and rating and remaining rows contains dataset values, and we will used above REVIEWS and RATINGS to train machine learning models. Below is the test data which contains only disease name and machine learning will predict Drug name and ratings.

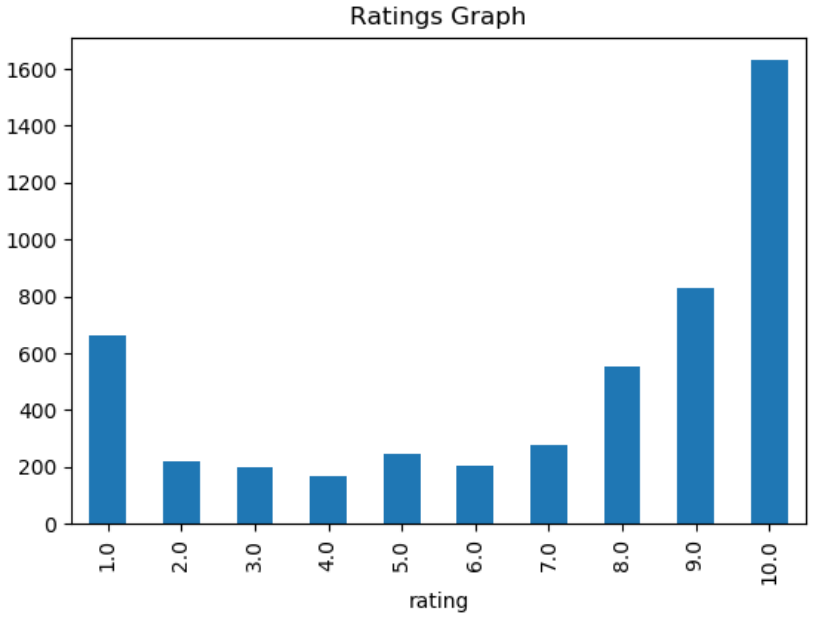

Figure 6. Drugs ratings graph.

In Figure 6, we can see dataset loaded and in graph x-axis represents ratings and y-axis represents total number of records which got that rating. Now close above graph and then click on 'Read & Pre-process Dataset' button to read all dataset values and then pre-process to remove stop words and special symbols and then form a features array.

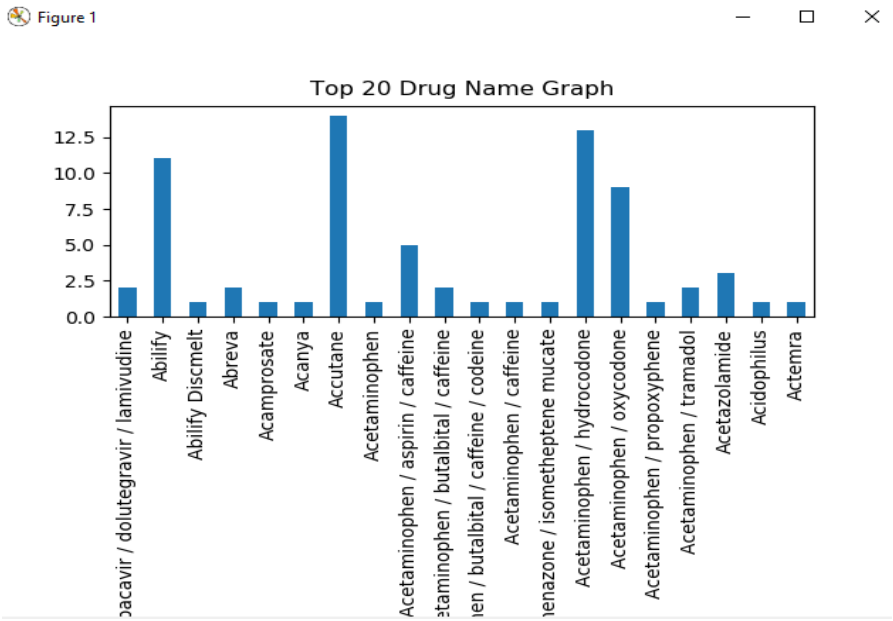

Figure 7. Drug names dataset.

In Figure 7, we can see from all reviews stop words and special symbols are removed and in graph I am displaying TOP 20 medicines exist in dataset. In above graph x-axis represents drug name and y-axis represents its count. Table 1 compares the performance of proposed method with existing methods. Here, Proposed TF-IDF and MLP resulted in superior Accuracy, Precision, Recall, and F1-SCORE as compared to existing Modified UCI. The graphical representation of table 1 is presented in figure 5. In Figure 9, for each disease name application has predicted recommended drug name and ratings.

Table 1. Performance comparison

Method	Precision	Recall	F1-Score	Accuracy
Existing Logistic regression	80.54	79.30	79.27	76
Existing SVC	70.51	71.18	70,46	67.80
Existing Ridge classifier	66.786	37.72	42.78	55.1
Existing Multimodal navie bayes	41.32	47.98	43.14	47.19
Existing SGDC	41.324	47.18	43.44	47.49
Proposed MLP	99.96	99.72	99.84	99.9

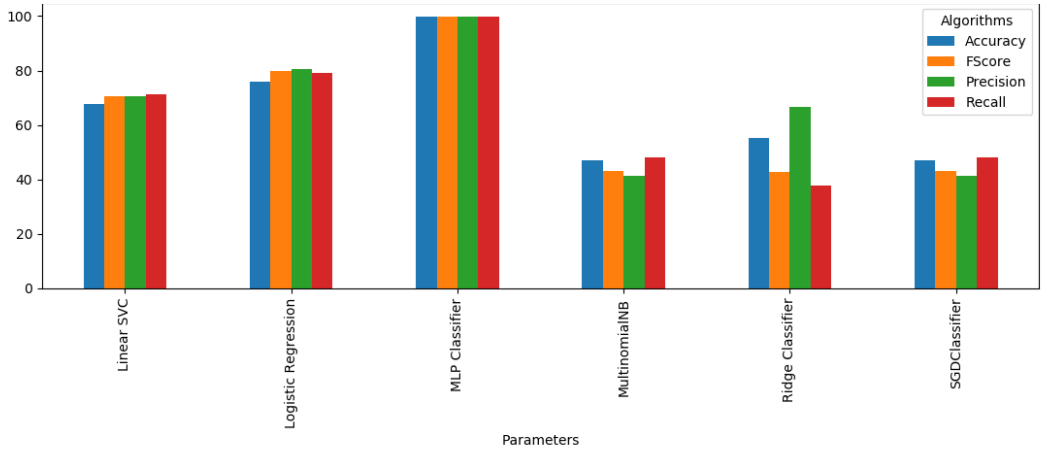

Figure 8. Performance comparison graph.

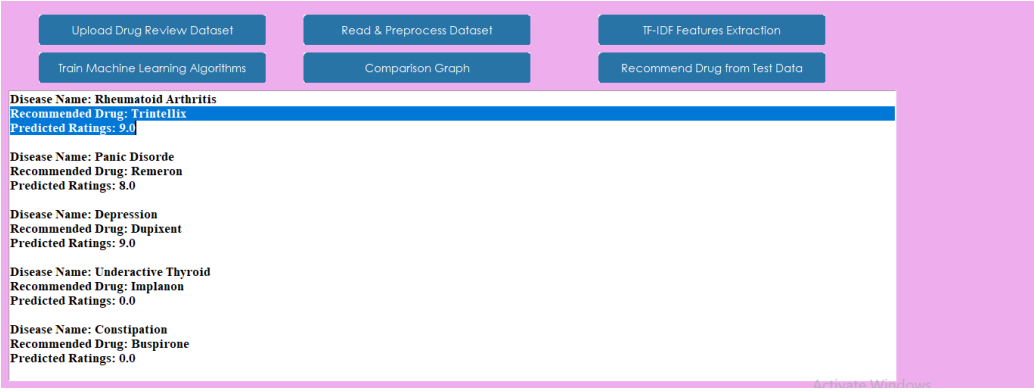

Figure 9. Drug recommendations from test data.

5. Conclusion

Reviews are becoming an integral part of our daily lives; whether go for shopping, purchase something online or go to some restaurant, we first check the reviews to make the right decisions. Motivated by this, in this research sentiment analysis of drug reviews was studied to build a recommender system using different types of machine learning classifiers, such as Logistic Regression, MLP, Multinomial Naive Bayes, Ridge classifier, Stochastic gradient descent, LinearSVC, applied on TF-IDF features. We evaluated them using five different metrics, precision, recall, f1score, accuracy, and AUC score, which reveal that the MLP on TF-IDF outperforms all other models with 99% accuracy. Future work involves comparison of different oversampling techniques, using different values of n-grams, and optimization of algorithms to improve the performance of the recommender system. Emergencies such as pandemics, floods, or cyclones can be helped by the medical recommender system. In the era of deep learning, recommender systems produce more accurate, quick, and reliable clinical predictions with minimal costs. As a result, these systems maintain better performance, integrity, and privacy of patient data in the decision-making process and provide precise information at any time.

References

[1] J. Ramos. "Using tf-idf to determine word relevance in document queries", in Proceedings of the first instructional conference on machinelearning, vol. 242, pp. 133–142, Piscataway, NJ, 2003

[2] K. Shimada, H. Takada, S. Mitsuyama, H. Ban, H. Matsuo, H. Otake, H. Kunishima, K. Kanemitsu and M. Kaku. "Drug-recommendation system for patients with infectious diseases". AMIA Annu Symp Proc. 2005; 2005:1112. PMID: 16779399; PMCID: PMC1560833.

[3] H. He, Y. Bai, E. A. Garcia and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning", 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), 2008, pp. 1322-1328, doi: 10.1109/IJCNN.2008.4633969.

[4] X. Lei, G. Anna Lisa, M. James and J. Iria. 2009. "Improving Patient Opinion Mining Through Multi-Step Classification. In Proceedings of the 12th International Conference on Text, Speech and Dialogue (TSD '09)". Springer-Verlag, Berlin, Heidelberg, 70–76. https://doi.org/10.1007/978-3-642-04208-9_13.

[5] A. Nikfarjam and G. H. Gonzalez. "Pattern mining for extraction of mentions of Adverse Drug Reactions from user comments". AMIA Annu Symp Proc. 2011; 2011:1019-26. Epub 2011 Oct 22. PMID: 22195162; PMCID: PMC3243273.

[6] C. Doulaverakis, G. Nikolaidis and A. Kleontas. "GalenOWL: Ontology-based drug recommendations discovery". J Biomed Semant 3, 14 (2012). https://doi.org/10.1186/2041-1480-3-14.

[7] L. Goeuriot, J. C. Na, W. Y. M. Kyaing, C. Khoo, Y. K. Chang, Y. L. Theng, and J. Kim. 2012. "Sentiment Lexicons for Healthrelated Opinion Mining. In Proceedings of the 2Nd ACM SIGHIT International Health Informatics Symposium (IHI '12)". ACM, New York, NY, USA, 219–226. https://doi.org/10.1145/2110363.2110390.

[8] R. N. Keers, S. D. Williams, J. Cooke and D. M. Ashcroft. "Causes of medication administration errors in hospitals: a systematic review of quantitative and qualitative evidence". Drug Saf. 2013 Nov;36(11):1045-67. doi: 10.1007/s40264-013-0090-2. PMID: 23975331; PMCID: PMC3824584.

[9] C. M. Wittich, C. M. Burkle and W. L. Lanier. "Medication errors: an overview for clinicians. Mayo Clin Proc. 2014 Aug;89(8):1116-25". doi: 10.1016/j.mayocp.2014.05.007. Epub 2014 Jun 27. PMID: 24981217.

[10] Y. Zhang, D. Zhang and M. M. Hassan. "CADRE: Cloud-Assisted Drug REcommendation Service for Online Pharmacies". Mobile Netw Appl 20, 348–355 (2015). https://doi.org/10.1007/s11036-014-0537-4.

[11] B. Danushka, M. Takanori and K. Kenichi. "Unsupervised Cross-Domain Word Representation Learning", 2015; arXiv:1505.07184

[12] A. Sarker, R. Ginn, A. Nikfarjam, K. O'Connor, K. Smith, J. Swetha, U. Tejaswi, G. Gonzalez. "Utilizing social media data for pharmacovigilance: A review, Journal of Biomedical Informatics", Vol. 54, 2015, pp. 202-212, no. 1532-0464, https://doi.org/10.1016/j.jbi.2015.02.004.

[13] A. Nikfarjam, A. Sarker, K. O'Connor, R. Ginn and G. Gonzalez. "Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features", J Am Med Inform Assoc. 2015 May;22(3):671-81. doi: 10.1093/jamia/ocu041. Epub 2015 Mar 9. PMID: 25755127; PMCID: PMC4457113.

[14] T. N. Tekade and M. Emmanuel, "Probabilistic aspect mining approach for interpretation and evaluation of drug reviews", 2016 International Conference on Signal Processing,

Communication, Power and Embedded System (SCOPES), 2016, pp. 1471-1476, doi: 10.1109/SCOPES.2016.7955684.

[15] L. Sun, C. Liu, C. Guo, H. Xiong, and Y. Xie. 2016. "Data-driven Automatic Treatment Regimen Development and Recommendation". In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, 1865–1874. DOI: https://doi.org/10.1145/2939672.2939866.

[16] J. Li, H. Xu, X. He, J. Deng and X. Sun, "Tweet modeling with LSTM recurrent neural networks for hashtag recommendation", 2016 International Joint Conference on Neural Networks (IJCNN), 2016, pp. 1570-1577, doi: 10.1109/IJCNN.2016.7727385.

[17] I. Korkontzelos, A. Nikfarjam, M. Shardlow, A. Sarker, S. Ananiadou, G. Gonzalez. "Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts", Journal of Biomedical Informatics, vol. 62, 2016, pp. 148-158, no. 1532-0464, https://doi.org/10.1016/j.jbi.2016.06.007.

[18] V. Gopalakrishnan and C. Ramaswamy. "Patient opinion mining to analyze drugs satisfaction using supervised learning, Journal of Applied Research and Technology", vol. 15, issue 4, 2017, pp. 311-319, no. 1665-6423, https://doi.org/10.1016/j.jart.2017.02.005.

[19] S. Alireza, K. Fatemeh and R. Nasrin. "Preventing the medication errors in hospitals: A qualitative study", International Journal of Africa Nursing Sciences, vol. 13, 2020, 100235, no. 2214-1391, https://doi.org/10.1016/j.ijans.2020.100235.

[20] G. Gurdin, J. A. Vargas, L. G. Maffey, A. L. Olex , N. A. Lewinski, S. T. McInnes. "Analysis of Inter-Domain and Cross-Domain Drug Review Polarity Classification". AMIA Jt Summits Transl Sci Proc. 2020 May 30; 2020:201-210. PMID: 32477639; PMCID: PMC7233089.

[21] S. Garg. "Drug Recommendation System based on Sentiment Analysis of Drug Reviews using Machine Learning", 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 175-181, doi: 10.1109/Confluence51648.2021.9377188.