IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Predictive Drug Recommendation Based On Patient Reviews And Disease Inputs

[1]Routhu Yeswanth Kumar, Student in Dept. Of Master of Computer Applications, at Miracle Educational Society Group of Institutions
[2]Mr. Ch. Kodandaramu, Associate Professor at Miracle Educational Society Group of Institutions
[3]Seera Sitalakshmi, Assistant Professor at Miracle Educational Society Group of Institutions

ABSTRACT

The emergence of exotic diseases points out a pressing void for swift medical aid. In this project, the authors focus on the recommendation of drugs utilizing a combination of patient reviews, disease description and machine learning aided sentiment analysis. The methodology incorporates TF-IDF, Bag of Words, and Word2Vec for feature extraction along with Logistic Regression and Multilayer Perceptron (MLP) as algorithms. For training and testing of the models the authors utilized the DRUGREVIEW dataset publicly available in UCI, reviews and drugs with ratings are used to anticipate drug outcomes. It was determined that for predicting drug outcomes MLP shows greater efficiency than the rest, hence it was chosen as the core algorithm in the developed system. Additionally doctors' prescribed medications are supplemented with suggested sentiments so that the chances of self medicine withdrawal are decreased and patients make better choices with regard to the medications.

Keywords: MLP, TF-IDF, Logistic Regression

INTRODUCTION:

Social media is now flooded with user generated content because of the growth of Web 2.0. This explosion has led to many studies on sentiment analysis, a subfield that deals with what people think and feel about a product, service, event, etc. simply put, analyzing reviews. Sentiment analysis or opinion mining can also be described as the process of determining the sentiment acerca of a text. For feedback, it is quite common to be tagged as positive or negative. However, for an accurate analysis such categories do not suffice and there is a need for more depth classifying what a user truly requires, for instance the level of satisfaction or the effectiveness.

In the particular area of medical check-up drug prescription recommended systems, the numerical sentiment score calibration rather than simple yes/no response system has shown adequate importance These challenges include overspecialization and data sparse aspects, scalability and even collaborative filtering CF, content based CB and hybrid forms of recommendation. Integrating user reviews' sentiment analysis into the recommendation model has been shown to yield better recommendations' predictions, a useful means in improving public health.

GAP IDENTIFIED BASED ON LITERATURE SURVEY:

Existing models doe based systems do not pay any attention to history and past experiences of the patients. Most existing models are more fixed on the generalized medical practice protocols. Furthermore, traditional ones hardly include the sentiment of the users particularly reviews of patients which would offer better suggestions on medications.

The other important gap is that most of the robust feature extraction models and their application to the analysis are employed and fully utilized. Some of the systems employ low level sequential, array or basic text analysis, TF-IDF and Word2Vec with MLP models are however not used in this area. In addition, most existing models concentrate on performance and hardly do optimization work which results in poor levels of accuracy and reliability of drug prediction.

This project fulfills these objectives by proposing a recommendation system that focuses on machine learning and drug review sentiment analysis. It performs feature extraction through TF-IDF and uses various high-performance algorithms for predictions. It uses data provided by patients, and thus, creates a bridge between the medical standards and how things are done in practice thereby improving the accuracy and trustworthiness of drug recommendations.

PROBLEM STATEMENT:

Ninety percent of patients with a new or complicated disease face issues in understanding which medicine would work for them and would end up causing horrendous health issues due to wrong medication.

Key Challenges:

1. Users' Emotions: The subjective cues in the reviews are tough to translate and thus can end in failure to add meaning to the insights.

2. Feature Engineering: The extraction of the most appropriate features that will cater for efficiency in machine learning models instilling accuracy.

3. Algorithm Classification: Machine learning algorithm that would give accurate and reliable predictions on drugs.

4. Disease and Drug Lifecycles: Different disease and drug data exist and should be data should be different disease and drug data and be robust and be thoroughly data driven.

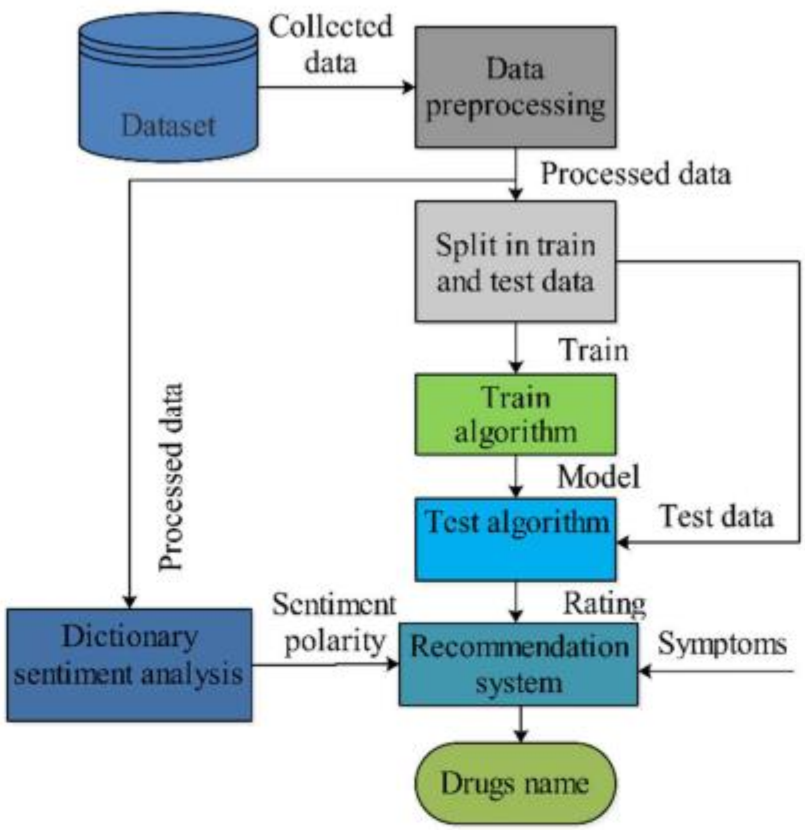
5. Solve Problems: The requirements in medical forecasting should meet the expectations of reality and what patients expect.

PROPOSED METHOD:

The suggested approach involves the combination of sentiment analysis and machine learning as the key elements of a trustworthy framework for drug recommendation. Preprocessing of the DRUGREVIEW data set is carried out in order to obtain the reviews, the ratings and the names of the drugs. Methods of feature extraction, such as TF-IDF, are used to convert text reviews into numbers. These features are used to train various prediction models such as drug reviews including Logistic Regression and MLP models.

Of those models, MLP has the greatest accuracy thus was used for the final solution. To predict drug ratings, the patients have to input the disease name and the system runs the suggested method to derive the cut-off drug and a sentiment-based rating. So within such comparative analysis of algorithms accuracy is shown and graphical outputs of performed models include model performance metrics. The part of sentiment analysis in conjunction with machine learning provides this method with drug recommendations that are correct and easy to use.

ARCHITECTURE:



DATASET:

This project is premised on the DRUGREVIEW dataset sourced from UCI repository. It has columns like drug name, condition, reviews, and ratings. The reviews and ratings act as sentiment information while conditions and drug names provide the basis for mapping disease-drug relationship. In this case, cleansing involves the deletion of stop words, patterns and irrelevant symbols, thus a clean dataset will be used to extract features. Textual reviews are transformed from the original form and represented as TF-IDF vectors that reflect word count and importance. These structured information are then used to train and test machine learning models allowing the models to accurately predict the name of drugs and the corresponding rating for a specific disease.

METHODOLOGY:

Data Preprocessing

In this stage, the DRUGREVIEW dataset which contains reviews, ratings, drug names and conditions is fetched.

The next step is to process the text so as to eliminate stop words, special symbols, and other non relevant fields.

The next step is to divide all the for the purpose of cross validation so that one part can be used to train and the second part to test the performance of the model.

Feature extraction

Train each model and implement TF-IDIF to turn text reviews into numerical vectors that indicate how often specific words were used and how important they are.

Employ Bag of Words and Word2Vec as other baseline methods for comparative feature extraction analysis.

Model Training:

Implement several machine learning techniques like Logistic Regression, Ridge Classifier, Naïve Bayes, and MLP.

Use TF-IDF vectors as features and drug names or ratings as labels for the task.

Use the following model evaluation strategies: accuracy, precision, recall and F1-score.

Performance Evaluation:

Rank algorithm performance to find out the best fit model in terms of accuracy.

Depict results with the help of comparative graphs where MLP performs better metrics.

Drug Recommendation:

Input disease names which are accepted from the test data.

Based on the prediction, select the drug that appears to be most effective and provide its rating based on sentiment.

Select drugs with high ratings for recommendations and make them credible for the patients.

Implementation and Testing:

Implement the system and EASY INTERFACE for dataset upload and TEST data upload.

Make predictions of diseases and drug recommendations with graphical representation on real time basis.

Graphical Visualization:

Plot graphs showing the distribution of dataset, the top drugs and performance metrics graphs.

Present the results of drug recommendation and predicted rating for them for clarity of understanding.

EVALUATION:

Precision:

Formula: $\text{Precision} = \frac{\text{True Positives}}{\text{True Positives}+\text{False Positives}}$

Recall (Sensitivity):

Formula: $\text{Recall} = \frac{\text{True Positives}}{\text{True Positives}+\text{False Negatives}}$

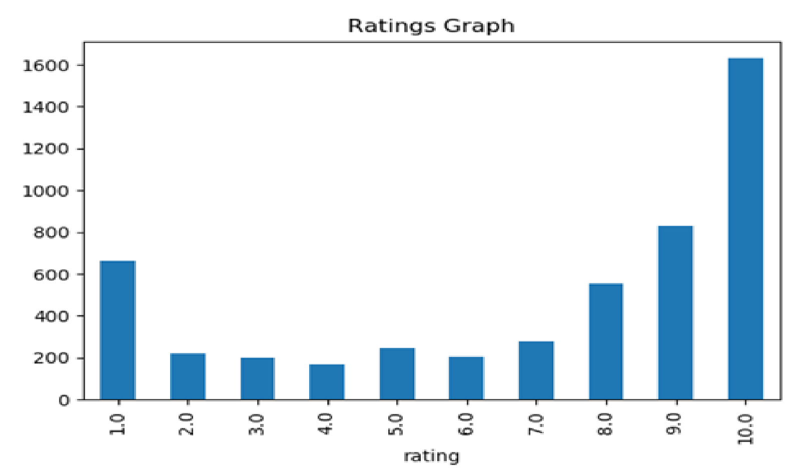
F1 Score:

Formula: $F1 = 2 \times \frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$

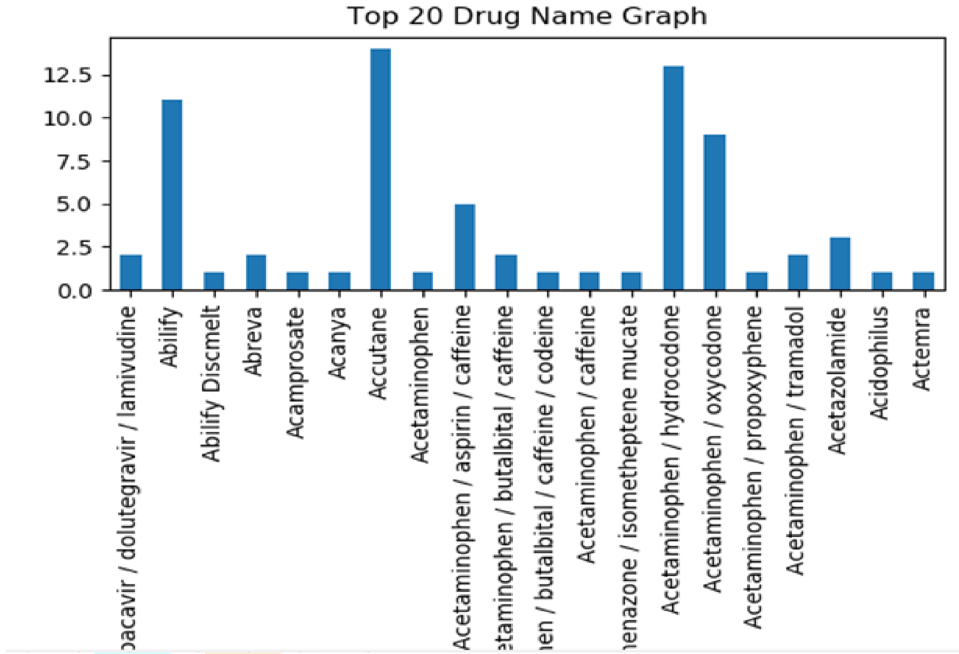
Accuracy:

Formula: $\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}}$

RESULTS:



x-axis represents ratings and y-axis represents total number of records which got that rating



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.

Logistic Regression Precision : 80.54683467044312
Logistic Regression Recall : 79.30203003457763
Logistic Regression F1-Score : 79.72928850771024
Logistic Regression Accuracy : 76.0

Linear SVC Precision : 70.51855963732795
Linear SVC Recall : 71.1815169715316
Linear SVC F1-Score : 70.46290195776827
Linear SVC Accuracy : 67.80000000000001

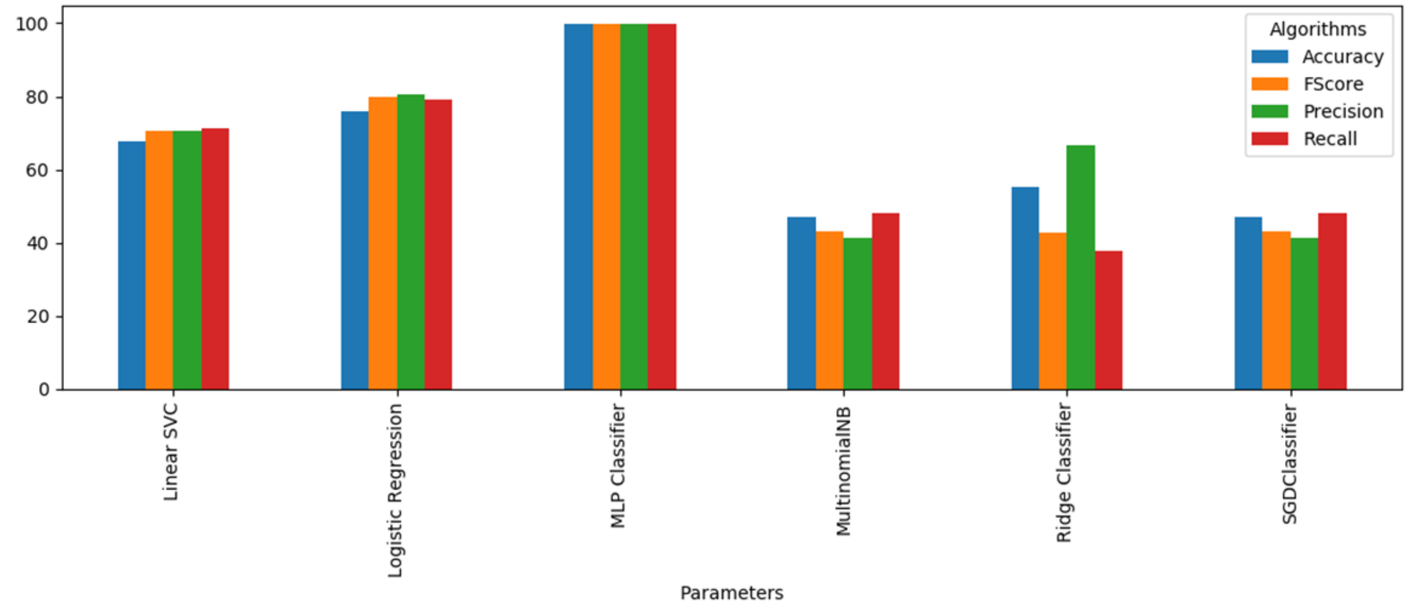
Ridge Classifier Recall : 37.72383152495436
Ridge Classifier F1-Score : 42.784203316921044
Ridge Classifier Accuracy : 55.1

Multinomial Naive Bayes Precision : 41.32611772759363
Multinomial Naive Bayes Recall : 47.984631801286945
Multinomial Naive Bayes F1-Score : 43.14147322230588
Multinomial Naive Bayes Accuracy : 47.199999999999996

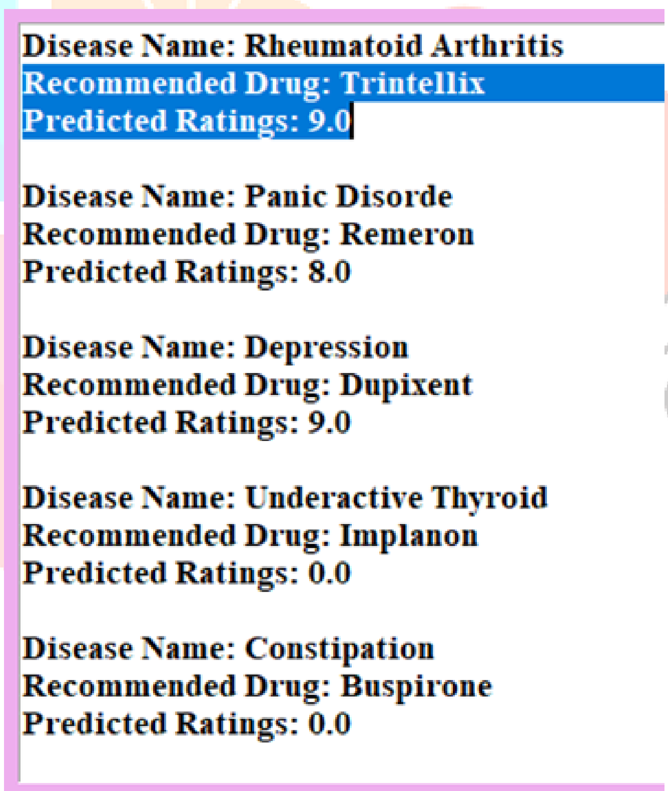
SGDClassifier Precision : 41.32611772759363
SGDClassifier Recall : 47.984631801286945
SGDClassifier F1-Score : 43.14147322230588
SGDClassifier Accuracy : 47.199999999999996

Multilayer Perceptron Classifier Precision : 99.96794871794872
Multilayer Perceptron Classifier Recall : 99.72222222222221
Multilayer Perceptron Classifier F1-Score : 99.84310356521149
Multilayer Perceptron Classifier Accuracy : 99.9

All algorithms MLP has got high performance



X-axis represents algorithm name and y-axis represents accuracy, precision recall and FSCORE where each different colour bar will represents one metric and in above graph we can see MLP got high performance.



Each disease name application has predicted recommended drug name and ratings

CONCLUSION

The aim of this system is to recommend drugs by the use of machine learning artificial intelligence patients. The integration of TF-IDF as a feature extraction and Multilayer Perceptron professional algorithms made sound and reliable predictions. DRUGREVIEW publicly available resource offers capabilities through which user sentiments and recommendations are connected bridging a missing link. The framework further strengthens the argument on bringing trustworthiness and wider applicability within medical spheres on the dataset by suggesting further work to be done on incorporating numerous diseases and live updates for realtime healthcare systems.

REFERENCES:

[1] B. Liu, Sentiment Analysis(Introduction and Survey) and Opinion Mining. 2012.

[2] X. Lei, X. Qian, and G. Zhao, "Rating Prediction Based on Social Sentiment from Textual Reviews," IEEE Trans. Multimed., vol. 18, no. 9, pp. 1910–1921, Sep. 2016, doi: 10.1109/TMM.2016.2575738.

[3] Y. Bao and X. Jiang, "An intelligent medicine recommender system framework," in Proceedings of the 2016 IEEE 11th Conference on Industrial Electronics and Applications, ICIEA 2016, Oct. 2016, pp. 1383–1388, doi: 10.1109/ICIEA.2016.7603801.

[4] R. Majethia, V. Mishra, A. Singhal, K. Lakshmi Manasa, K. Sahiti, and V. Nandwani, "PeopleSave: Recommending effective drugs through web crowdsourcing," in 2016 8th International Conference on Communication Systems and Networks, COMSNETS 2016, Mar. 2016, doi: 10.1109/COMSNETS.2016.7440000.

[5] R. C. Chen, Y. H. Huang, C. T. Bau, and S. M. Chen, "A recommendation system based on domain ontology and SWRL for anti-diabetic drugs selection," Expert Syst. Appl., vol. 39, no. 4, pp. 3995–4006, Mar. 2012, doi: 10.1016/j.eswa.2011.09.061.

[6] J.-C. Na and W. Y. M. Kyaing, "Sentiment Analysis of User-Generated Content on Drug Review Websites," J. Inf. Sci. Theory Pract., vol. 3, no. 1, pp. 6–23, Mar. 2015, doi: 10.1633/jistap.2015.3.1.1.

[7] M. E. Basiri, M. Abdar, M. A. Cifci, S. Nemati, and U. R. Acharya, "A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques," Knowledge-Based Syst., vol. 198, p. 105949, Jun. 2020, doi: 10.1016/j.knosys.2020.105949.

[8] S. Vijayaraghavan and D. Basu, "Sentiment Analysis in Drug Reviews using Supervised Machine Learning Algorithms," arXiv, Mar. 2020, Accessed: Nov. 20, 2020. [Online]. Available: http://arxiv.org/abs/2003.11643.

[9] V. Doma et al., "Automated Drug Suggestion Using Machine Learning," in Advances in Intelligent Systems and Computing, Mar. 2020, vol. 1130 AISC, pp. 571–589, doi: 10.1007/978-3-030-39442-4_42.

[10] A. A. Hamed, R. Roose, M. Branicki, and A. Rubin, "TRecs: Time-aware twitter-based drug recommender system," in Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2012, 2012, pp. 1027– 1031, doi: 10.1109/ASONAM.2012.178.

[11] C. Chen, L. Zhang, X. Fan, Y. Wang, C. Xu, and R. Liu, "A epilepsy drug recommendation system by implicit feedback and crossing recommendation," in Proceedings - 2018 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovations, SmartWorld/UIC/ATC/ScalCom/CBDCo, 2018, doi: 10.1109/SmartWorld.2018.00197.

[12] D. Chen, D. Jin, T. T. Goh, N. Li, and L. Wei, "ContextAwareness Based Personalized Recommendation of AntiHypertension Drugs," J. Med. Syst., vol. 40, no. 9, pp. 1– 10, Sep. 2016, doi: 10.1007/s10916-016-0560-z.

[13] A. Gottlieb, G. Y. Stein, E. Ruppin, R. B. Altman, and R. Sharan, "A method for inferring medical diagnoses from patient similarities," BMC Med., vol. 11, no. 1, p. 194, Sep. 2013, doi: 10.1186/1741-7015-11-194.

[14] K. Shimada, K. Fujikawa, K. Yahara, and T. Nakamura, "Antioxidative Properties of Xanthan on the Autoxidation of Soybean Oil in Cyclodextrin Emulsion," 1992. Accessed: Jul. 29, 2020. [Online]. Available: https://pubs.acs.org/sharingguidelines.

[15] Q. Zhang, G. Zhang, J. Lu, and D. Wu, "A framework of hybrid recommender system for personalized clinical prescription," in Proceedings - The 2015 10th International Conference on Intelligent Systems and Knowledge Engineering, ISKE 2015, Jan. 2016, pp. 189–195, doi: 10.1109/ISKE.2015.98.