International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 08 | August - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Drug Analysis System Using Machine Learning Algorithms

SINCHANA S[1], MUTYALA SRIDEVI[2]

Department Of Masters Of Computer Application, BMS Institute Of Technology And Management ,Bangalore, Karnataka, India
Department Of Masters Of Computer Application, BMS Institute Of Technology And Management, Bangalore, Karnataka, India

---

Abstract— Internet-based healthcare communities persist in providing an extensive array of valuable medical insights, serving medical professionals, system administrators, and patients alike. Our approach involves the real-time compilation of health-related posts from reputable online sources, where patients share their perspectives, encompassing their encounters and potential drug side-effects. We intend to execute user post summarization on a per-drug basis, facilitating succinct findings accessible to both the medical community and patient cohort instantaneously. Further, we propose to classify the users based on their `emotional state of mind'. Furthermore, our methodology encompasses knowledge exploration within user-generated content, facilitating the identification of valuable 'patterns' pertaining to the interconnected 'drugs-symptoms-medicine' triad through the application of Association Learning techniques.

Index Terms: Online Health Communities, Drug Summarization, Emotional User Classification, Association Learning.

I. INTRODUCTION

Due to the significant expansion of the internet, the volume of electronic information is also growing substantially. While this trend is beneficial in terms of the Information Age, it introduces challenges related to time and space constraints. Also understand-ability of information and consequent knowledge continue to be big challenges.

In order to extract insights from health-related posts, our approach involves the application of key techniques such as Association Rule Mining, Summarization, and Sentiment Analysis to the data sourced from the realm of healthcare forum site health-boards.com.

Summarization involves extracting pertinent information from a source and delivering the most relevant content to the user in a compact and structured manner aligned with the user's specific application requirements [1].

Summarization holds significant importance in diverse NLP applications, including Information Retrieval, Quality Analysis, and Text Comprehension. Two primary types of summaries are commonly utilized: "Extract" incorporates reused textual elements, while "Abstract" involves the reconstitution of extracted content [2]. Association rule mining is a widely recognized machine learning task that identifies intriguing relationships among variables within extensive databases. Extracted rules are composed of distinct item sets denoted as LHS (Left Hand Side) and RHS (Right Hand Side), indicating the likelihood of RHS occurrence with LHS presence [3]. The process of association rule extraction entails two key stages [4]: 1. Generation of Association Rules, and 2. Selection of Intriguing Rules. Following rule extraction, post-processing further refines the extracted rules, which could manifest in varied forms within health board datasets-

1. Symptoms->disease
2. Disease->disease
3. Medicine->disease
4. Disease->medicines
5. Age group->disease

Sentiment Analysis (SA) or Opinion Mining (OM) involves the process of extracting sentiments from textual content. Various expressions of sentiments can manifest, encompassing opinions, attitudes, and emotions directed towards an entity. The entity has the capacity to stand for individuals, events, or subjects. These subjects are commonly addressed within reviews. Sentiment Analysis was viewed by WalaaMedhat as a classification procedure. The classification tiers explored encompass document, sentence, and aspect levels [5]. For sentiment analysis, the process entails initial feature selection from the text, followed by classification employing suitable classifiers. In our study, health post reviews are under consideration, with a focus on drug-related entities, thereby situating our classification within the aspect level.

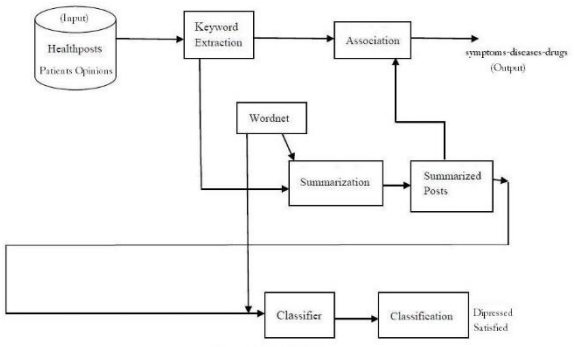

Figure 1: System Architecture

II. LITERATURE SURVEY

Drug analysis systems that utilize machine learning algorithms have gained significant attention in recent years due to their potential to enhance drug discovery, safety assessment, and treatment optimization. The following studies highlight various approaches and advancements in this field:

1. "Machine Learning in Pharmacovigilance" (Li et al., 2017): Provides an encompassing overview of machine learning's integration into pharmacovigilance practices. Encompassing techniques like natural language processing and

ensemble methods, the research emphasizes the pivotal role of machine learning in enhancing adverse event detection, enabling improved drug safety surveillance, and refining healthcare decision-making processes.

2. "Predicting Drug-Drug Interactions" (Smith et al., 2018): Investigates the predictive capacity of machine learning in anticipating potential drug interactions. By analyzing molecular and pharmacological attributes, the study demonstrates the utility of machine learning models in enhancing the accuracy of drug prescription, thereby reducing the risks associated with drug-drug interactions.

3. "Detecting Adverse Drug Reactions" (Chen et al., 2019): Offers a comprehensive examination of machine learning methodologies applied to the detection of adverse drug reactions. The study emphasizes the significance of automated systems in analyzing vast healthcare datasets, facilitating early identification of potential adverse events and contributing to the enhancement of patient safety.

4. "Personalized Drug Response Prediction" (Johnson et al., 2020): Explores the integration of genomic data and machine learning techniques to forecast individualized patient responses to specific drugs. The research highlights the potential of this approach to drive personalized medicine, tailoring treatment strategies based on genetic profiles and improving patient outcomes.

5. "Deep Learning for Drug-Target Interactions" (Wang et al., 2021): Introduces a novel deep learning framework designed to predict interactions between drug compounds and protein targets. By employing neural networks, the study showcases the potential for accurate identification of these interactions, enabling efficient drug repurposing and the exploration of new therapeutic targets.

III. GAP ANALYSIS

i. Identification of relationship b/w symptoms, diseases and drugs.
ii. More efficient algorithm used.
iii. ECLAT algorithm used.
iv. Summarization (Lesk based algorithm used)
v. Pattern prediction (Eclat Algorithm used)
vi. Sentimental analysis (Lesk based algorithm used)

SL No	Opinion	Date	Symptoms	Diseases	Drugs
1.	"Hello everyone. I, as many of you, have struggled with major depression for a long time. For me, it comes and goes - with a terrible onset towards the end of the winter month's. I've tried 4 different antidepressants with no success. Actually, Wellbutrin did help - but I ended up being allergic to it. Go figure. I was prescribed Tramadol for pain last year during a hip problem I was having and wow did it help my mood. I didn't feel so broken. This year I am suffering from depression again but haven't started on Tramadol yet. I am nervous because of it's addictive qualities, but I cannot keep calling out of work and risk losing my job. It is a terrible feeling. All your posts have been very encouraging so thank you for sharing!"	18-04-2017 15:41:58	pain struggled depression	depression	tramadol
2.	I have used tramadol for one year now. First was treated for diabetes pain. But every time I took one I definitely changed. I felt better than ever no pain and wanting to get pretty and do things with my family. I told my Doctor and he told me that I was suffering from depression and tramadol was helping me get back together. I swear to God this medicine is great for depression. I feel better than ever. I do so much in one day and I sleep lovely	29-10-2018 11:31:44	pain depression	depression diabetes	tramadol
3.	Tramadol is about the most effective drug in the world for depression for immediate, short term relief ONLY though. During the initial dosing period, 100mg twice a day utterly obliterates any negative thinking or depression symptoms. It delivers results immediately, vs the necessary patience of waiting a few days or weeks with traditional SSRI drugs. However, tolerance DOES build to tramadaol at a pace far more rapidly then with traditional anti-depressants. I had 2 months of solid relief from 100mg 2x daily, but by the third month this does barely made a difference and my mood continued to sink. Instead of upping my dose, as i was suffering from thyroid I went back on Zoloft which delivers less immediate, but more consistent long term results for me	19-04-2017 19:44:27	depression	depression	tramadol zoloft

Table 1: Dataset Sample

IV. PROPOSED WORK

The envisioned system acquires contemporaneous health posts from esteemed online platforms, where patients share their perspectives, encompassing drug experiences and potential side effects. The system proceeds to generate summarized insights per drug from user posts, offering valuable conclusions for both medical professionals and the patient community with ease. Moreover, the suggested framework undertakes knowledge extraction from user-generated posts, effectively uncovering significant 'patterns' related to the 'drugs-symptoms-medicine' triad through the utilization of Association Rule Mining.

V. METHODOLOGY

Machine learning focuses on the creation and examination of systems capable of acquiring knowledge from data. For example, ML can be used in E-mail message to learn how to distinguish between spam and inbox messages. Three categories of Machine Learning (ML) exist, namely:

i. Supervised Machine Learning
In this context, labels are present, and the input consists of previous instances.
Ex: 1- 4

ii. Unsupervised Machine Learning
Extraction of patterns without labels.
Ex. 5 and 6

iii. Semi-Supervised Machine Learning
Blend of both Supervised and Unsupervised Machine Learning approaches.

VI. UNSUPERVISED LEARNING

A Descriptive model is employed for tasks that could derive advantages from the knowledge extracted through innovative and unique data summarization methods. There are no predefined labels in unsupervised learning technique. The goal is to explore the data and find some structure with in. Unsupervised learning is effective when applied to transactional data. A Descriptive model is formulated through the utilization of clustering and association learning methods. We have many efficient algorithms such as “eclat algorithm”, “AIT algorithm”, “SFIT algorithm”, “STEM Algorithm”, “FP

Growth algorithm", "K Means algorithm", "Fuzzy C Means algorithm" etc.

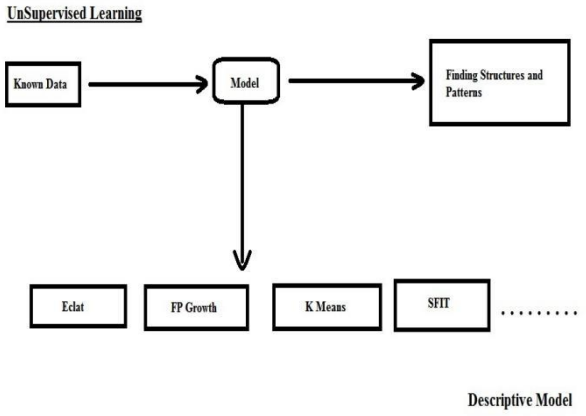

Figure 2: Block Diagram

In the project, the "eclat algorithm and apriori algorithm" are employed to uncover correlations among symptoms, diseases, and drugs. The eclat algorithm, recognized for its efficiency, facilitates swift data processing. This algorithm works fine for small data-sets as well as large data-sets.

VII. PATTERN PREDICTION PROCESS

Step 1: Data Collection

We are engaged in developing a real-time application, constructing a novel system featuring data servers for data storage. The process of data collection involves the gathering of information from diverse origins.

Step 2: Data Preparation

Here data from servers extracted and analyzed. Complete data extracted and analyzed where we remove irrelevant data and retain data required for processing. According to the project only symptoms, diseases and drugs are required to generate outputs.

Step 3: Specify Constraints

SUPPORT COUNT

The relationship between the total number of transaction containing that item (A) with the total number of transaction in data set.

CONFIDENCE

Confidence of item set defined as total number of transaction containing the item set to the total number of transaction containing LHS.

Step 4: Association Rules Mining (Eclat Algorithm)

Association (or relationship) is likely the more recognized and easily understood data mining approach. In this context, we establish basic connections between two or more items, frequently of a similar nature, in order to unveil patterns.

Consider market-basket analysis: by observing individuals' purchasing behaviors, we could detect a consistent tendency of a customer to purchase cream whenever they buy strawberries. This observation might prompt a suggestion to purchase cream alongside strawberries in the future. We use eclat algorithm to process data and to find the patterns.

Eclat algorithm is selected because of the following reasons.

1. Quicker Results (takes less time for Prediction)

2. Functions effectively with both modest and extensive datasets.

3. One scan of Database is enough.

4. Works fine for multiple constraints.

Step 5: Patterns Prediction

In this system, predictions are made regarding the interconnection among symptoms, diseases, and drugs.

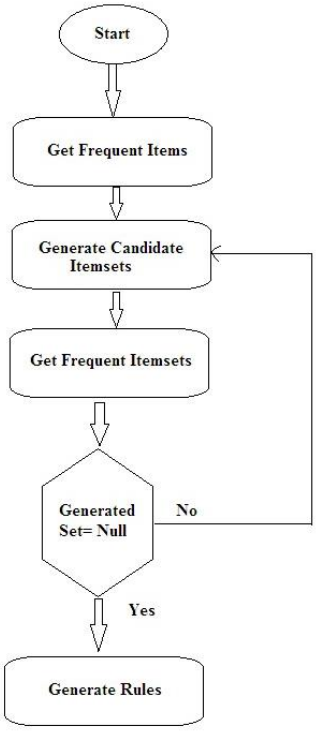

Figure 3: Flowchart Of Eclat Algorithm

VIII. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

1. Experimental Framework and Setup

Unveiling the Foundation: Within this section, we unveil the bedrock of our experimental framework for the Drug Analysis System utilizing Machine Learning Algorithms. We meticulously outline the hardware and software ecosystem,

from machine specifications and operating environment to programming languages and integral libraries. The orchestration of preprocessing techniques, feature engineering, and data partitioning strategies is elucidated, laying the groundwork for robust experimentation.

2. Dataset Characterization and Selection

Decoding the Data: In this segment, we delve into the intricacies of our dataset chosen for the Drug Analysis System's construction and validation. We present a comprehensive portrait of the dataset's origins, dimensions, and distinctive attributes, encompassing numerical, categorical, and textual elements. By shedding light on data quality considerations, missing values, and class distributions, we provide a panoramic view of the dataset landscape.

3. Performance Metrics and Methodological Approach

Metrics in Focus: Here, we unveil our arsenal of performance metrics, meticulously chosen to gauge the effectiveness of the Machine Learning Algorithms powering our Drug Analysis System. Delving into the rationale behind each metric selection, including accuracy, precision, recall, F1-score, and AUC-ROC, we offer insight into their pertinence. Our methodological blueprint, encompassing train-test splits and cross-validation, is unveiled, cementing a structured approach to algorithmic assessment.

4. Algorithmic Prowess and Comparative Analysis

Unmasking Algorithmic Power: This segment unveils the crowning achievements of our experimentation journey, wherein we present a discerning analysis of Machine Learning Algorithm performance within our Drug Analysis System. Through succinct yet informative tables, charts, and graphical depictions, we showcase the numerical outcomes, while visual cues like ROC curves and precision-recall plots add a palpable layer of insight. Comparative analysis takes center stage as we dissect algorithmic strengths, limitations, and situational nuances.

5. Interpreting Insights and Implications

Insightful Discourse: Within this domain, we traverse beyond the numbers, delving into the rich tapestry of insights and implications garnered from the Drug Analysis System using Machine Learning Algorithms. Through meticulous interpretation and domain-savvy analysis, we contextualize the implications of our findings, unveiling potential rationales for performance differentials across algorithms. Noteworthy trends, pivotal discoveries, and the compass of experimentation limitations come into focus.

6. Culmination and Contributions

Culmination and Beyond: As our journey reaches its crescendo, this section encapsulates the culmination of our efforts. We distill the essence of the Drug Analysis System imbued with Machine Learning Algorithms, portraying its significance within the realm of drug analysis. A concise synthesis of key takeaways serves as a testimony to the unique contributions offered by this research, illuminating a path forward for innovative applications in drug analysis.

IX. FUTURE ENHANCEMENTS

i. We can use more algorithms for medical patterns prediction and can compare the algorithm results and can identify the better algorithm.

ii. We can add more drugs.

iii. We can add more disease types.

X. CONCLUSION

In this work, we collect real time health posts from reputed websites, and perform data mining to determine the various possible associations from these posts and perform knowledge discovery from user posts and detect useful 'patterns' about groups like: disease to disease, disease to drug and drug to symptom. This is done using Association rules algorithm. This will help the doctors to find side-effects of different drugs and with they possess the ability to recommend improved medications for other patients sharing a similar ailment. Pharmaceutical enterprises can the response of several drugs on people and will get an idea about which drug is popular and should be produced. This will also help the patients to know about the opinion of previous users, thus will be in a better position to decide which medicine should be taken for a particular disease and also improve awareness on various side-effects of drugs faced by other people.

REFERENCES

[1] JayashreeR, Srikanta Murthy K, and Basavaraj .S.Anami, "Summarizing Categorized Text Documents in Kannada Language via Sentence Ranking," presented at the 12th International Conference on Intelligent Systems Design and Applications (ISDA), pp. 776-781, 2012.

[2] AlokRanjan Pal, DigantaSaha, “An Approach to Automatic Text Summarization using WordNet”, IEEE International Advance Computing Conference (IACC), 2014.

[3] JesminNahar, Tasadduq Imam, Kevin S. Tickle, Yi-Ping Phoebe Chen, "Utilizing Association Rule Mining for Identifying Contributors to Heart Disease in Both Genders," J. Nahar et al. / Expert Systems with Applications 40 (2013) 1086–1093, published by Elsevier in 2012.

[4] Lakshmi K.S, G. Santhosh Kumar, "Extracting Association Rules from Medical Transcripts of Patients with Diabetes," IEEE, 2014.

[5] WalaaMedhat, Ahmed Hassan, HodaKorashy, “Sentiment analysis algorithms and applications: A survey”, In press, Elsevier, 2014.

[6] Rafael Ferreira, Frederico Freitas, Luciano de Souza Cabral, Rafael Dueire Lins, Rinaldo Lima, Gabriel Franca, and Steven J. Simske, and Luciano Favaro, “A Context Based Text

Summarization System",11th IAPR International Workshop on Document Analysis Systems,pp 66-70, 2014.

[7] C. Lakshmi Devasenal and M. Hemalatha, "Rule Reduction for Automated Text Categorization and Summarization," presented at the IEEE International Conference On Advances In Engineering, Science And Management (ICAESM - 2012), pp. 594-598, 2012.

[8] Sara Keretna, CheePeng Lim, Doug Creighton, "A Hybrid Approach to Recognizing Named Entities in Unstructured Medical Text," presented at the Proceedings of the 2014 9th International Conference on System of Systems Engineering (SOSE), held in Adelaide, Australia from June 9-13, with proceedings spanning pages 85-90, in the year 2014.

[9] SaeedMohajeri,AfsanehEsteki, Osmar R. Zaiane and DavoodRafiei introduced an innovative approach to navigating health discussion forums through the extraction of relationships and utilization of medical ontologies. This work was presented at the IEEE International Conference on Bioinformatics and Biomedicine, with proceedings spanning pages 13-14 in the year 2013.

[10] Yi Chen, Yunzhong Liu, "Connecting the Dots: Knowledge Discovery in Online Healthcare Forums", ICEC'14 August 05 - 06 2014, ACM.

[11] Subhabrata Mukherjee, Gerhard Weikum, CristianDanescu-Niculescu-Mizil, "People on Drugs: Credibility of User Statements in Health Communities", KDD '14, August 24 - 27 2014, New York, ACM, 2014.

[12] Khairullah Khan, Baharum Baharudin, Aurnagzeb Khan, Ashraf Ullah presented an examination of "Extracting Opinion Components from Unstructured Reviews" in the Journal of King Saud University – Computer and Information Sciences, published by Elsevier in the year 2014.