

Data Analytics in the Pharmacology Domain

Maryam Qusay Yousif Helae, Trent University, Canada

Dariusz Ebrahimi, Thompson Rivers University, Canada

Fadi Alzhouri, Trent University, Canada*

ABSTRACT

Data mining approaches such as natural language processing play a fundamental role in the healthcare sector and, exclusively, the pharmacology domain. The substantial feedback and experiences shared by the patients on different drugs are employed to perform opinion mining on the reviews, which will help the decision-makers to improve the medications' quality and provide the optimal medical outcomes. Based on that, the drug review data set from the UCI machine learning repository is used. The objective of this study is to conduct a sentiment analysis of the patients' reviews to obtain their satisfaction with different drugs using the random forest (RF) machine learning model. In addition, finding out the best drugs for different conditions based on patients' reviews is done by implementing the long short-term memory. Finally, the authors predict the patients' medical conditions based on their reviews by performing the support vector machine and RF classifiers. The knowledge of the patients' medical condition and satisfaction will lead to a noticeable improvement in the pharmaceutical and medical consequences.

KEYWORDS

Big Data, BoW, Data Mining, Drug Reviews, Healthcare, LSTM, Machine Learning, ML, RF, Sentiment Analysis, Social System, SVM, TF-IDF

1. INTRODUCTION

Big data analysis plays a significant role in different real-life domains such as politics, business, and health care (Ahmed, 2017). Furthermore, analytic technology provided a salient contribution to the health care system. For instance, the recruitment of machine learning techniques and deep learning approaches enhanced the development of disease classification, early cancer detection, surgery improvement, and drug discovery and refinement (Islam et al., 2018). Because of the role of data analysis within the healthcare sector, the pharmaceutical domain will be the scope of this paper.

There are two main surveillance systems that are used for monitoring the safety and efficacy of the marketed drugs. The first is Vaccine Adverse Event Reporting System (VAERS), which was implemented by the United States Food and Drug Administration (Shimabukuro et al., 2015). The second is the Yellow Card scheme created by the United Kingdom Medicines and Healthcare Products Regulatory Agency (O'Donovan et al., 2019). In fact, there are many limitations of the clinical trials because of specific test protocols, for instance, the limited number of participants compared to the

DOI: 10.4018/ijbdah.314229

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

real population, the standardized experiment conditions, and the determined period in which the study is conducted. Thus, it is not very efficient to generalize pharmaceutical product safety results; this highlights the importance of pharmacovigilance to counter this issue (Gräßer et al., 2018).

Recently, a new methodology was constructed to overcome the limitation of a medication's clinical trials and to generalize a drug's efficiency within a bigger population. Accordingly, information technology (De Smedt & Daelemans, 2012), such as social media, online drug review applications, and websites, was employed to gather patient reviews about the drugs. The reviews include drug side effects, patient satisfaction rates, and patient experiences (Gräßer et al., 2018). Further, the online drugs reviews technique provides a plethora of useful information that contributes to the pharmacology domain, which helps to optimize drugs, provide optimal medical outcomes, and improve drug marketing and sales (Liu et al., 2020). It is therefore vital to scrape the data from the websites (Tenorio de Farias et al., 2021) and apply efficient analysis techniques to it; herein lies the importance of big data analysis models such as machine learning and deep learning.

The problem statement of this paper focuses on, first, conducting a sentiment analysis of patient reviews to determine satisfaction, whether positive, negative, or neutral, with different drugs using machine learning models. Second, finding out the best drug ratings (1–10 stars) for different conditions based on patient reviews using deep learning. Third, predicting patients' medical conditions based on their reviews.

Knowledge of patients' medical conditions and the drug recommendations ratings will help the patients to choose better medicines, especially when medical advice resources are limited and healthcare systems are overwhelmed, such as during the COVID-19 pandemic (Zeroual et al., 2020). Moreover, learning about patient satisfaction with the drugs will provide feedback to pharmaceutical companies to achieve better medical outcomes. Further, demonstrating the importance of the problem statement highlights the contributions of this study to the pharmacology domain. In addition to that, predicting patients' medical conditions based on their descriptions is a recent contribution to the research domain.

2. RELATED WORK

Natural language processing and sentiment analysis have increasingly drawn interest in the healthcare domain (Abirami & Askarunisa, 2017). Exclusively, the pharmacology sector, where some of the researchers investigated in drug reviews analysis. Online resources such as Twitter and drug review applications enable patients to share feedback about their medications and check others' reviews (Gopalakrishnan & Ramaswamy, 2017). Moreover, patient feedback will help doctors make better decisions about prescriptions and improve drug quality (Youbi & Settouti, 2021).

Researchers have focused on developing different medical lexicons to improve the benchmark of the sentiment analysis results. For instance, Asghar et al. (2016) employed pointwise mutual information (PMI), term frequency (TF), and inverse document frequency (IDF) to obtain the polarity score of the SentiWordNet lexicon (SWN) and health-related words using Web Lexicon (WL) to improve and develop medical lexicon quality. In addition to that, Liu and Lee (2019) implemented word embedding techniques in the SWN lexicon, which led to new medical sentiment phrases being added to the SWN lexicon. Further, the efficiency of the comprehensive SWN lexicon was tested by implementing the sentiment analysis on the drug reviews data using position encoding for feature extraction within the radial basis function network (RBFN), support vector machine (SVM), random forest (RF), and naive Bayes (NB) models for classifying patient emotions.

The study accomplished by Youbi and Settouti (2021) analyzed patients' feelings about drugs by using machine learning and deep learning techniques. Further, the VADER sentiment analysis algorithm was performed to extract sentiments from the data. In addition, feature extraction techniques such as the bag of words (BoW), TF-IDF, and *N*-gram approaches were applied to the data. Finally, SVM, RF, multinomial naive Bayes (MNB), convolutional neural network (CNN), recurrent neural

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/data-analytics-in-the-pharmacology-domain/314229

Related Content

Medical Device Security

Md Abdullah Al Momin (2022). *Security, Data Analytics, and Energy-Aware Solutions in the IoT* (pp. 173-191).

www.irma-international.org/chapter/medical-device-security/295908

Data Analytics in the Pharmacology Domain

Maryam Qusay Yousif Helae, Dariush Ebrahimiand Fadi Alzhouri (2022). *International Journal of Big Data and Analytics in Healthcare* (pp. 1-16).

www.irma-international.org/article/data-analytics-in-the-pharmacology-domain/314229

Data Collection and Analyses Applying Unmanned Helicopter (UAV) Remote Sensing to Survey Water Chestnut Invasive Species

Tao Tang, Chenliuli Jiangand Mary Perrelli (2020). *International Journal of Data Analytics* (pp. 38-51).

www.irma-international.org/article/data-collection-and-analyses-applying-unmanned-helicopter-uav-remote-sensing-to-survey-water-chestnut-invasive-species/244168

Smart Healthcare Apps for Quality Cancer Patient Support

Angelina Kouroubali, Lefteris Koumakis, Haridimos Kondylakisand Dimitrios G. Katehakis (2020). *International Journal of Big Data and Analytics in Healthcare* (pp. 28-48).

www.irma-international.org/article/smart-healthcare-apps-for-quality-cancer-patient-support/253844

A Survey on Prediction Using Big Data Analytics

M. Supriyaand A.J. Deepa (2017). *International Journal of Big Data and Analytics in Healthcare* (pp. 1-15).

www.irma-international.org/article/a-survey-on-prediction-using-big-data-analytics/197438