International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

AN INTELLIGENT DISEASE PREDICTION AND DRUG RECOMMENDATION PROTOTYPE BY USING MULTIPLE APPROACHES OF MACHINE LEARNING ALGORITHMS

Amtul Noor Afza[1], Anees Fatima[2], G Usha Kiran[3], Ayesha Sarwath[3]
[1,2,3]UG Scholar, Department of CSE, Shadan Women's College of Engineering and Technology, Hyderabad,
[4]Asst Professor, Department of CSE, Shadan Women's College of Engineering and Technology

Received: 05-04-2023 Accepted: 11-05-2023 Published: 18-05-2023

ABSTRACT:
Large blocks of data must be analyzed and explored by utilizing the data mining procedures in order to uncover significant patterns and trends. Medical databases are one area where the data mining procedures can be utilized. Many people all over the world are struggling with their health and medical diagnoses. Massive amounts of data are produced by hospital information systems (HIS), yet it might be difficult to extract knowledge from diagnosis case data. By just giving the symptoms they are experiencing, patients can quickly learn about the sickness they are experiencing and the medication that can assist, treat it using the approaches utilized in this paper. In this paper, we give drug recommendations relied on ratings and conditions to customers. To analyze the reviews and finally, probabilistic and weighted average methodologies are utilized to recommend the medications. Each model and strategy utilized in this paper is described in detail. The experimental findings presented in this work can be utilized in future studies and for a variety of different medicinal applications.

1. INTRODUCTION

A recommender prototype, broadly defined, is a prototype that anticipates the ratings a customer would give to a particular item. The customer will subsequently be given a ranking of these forecasts. Several household names including Google, Instagram, Spotify, Amazon, Reddit, Netflix, etc. employ them. Relied on the customer's profile, a recommender prototype can determine if a specific customer will favour an item or not. Both the service providers and customers can benefit from recommender prototype. They lower the transaction costs associated with locating and choosing products in an online buying setting. The utilization of recommender prototype is wide spread; with well-known examples include medicine recommenders, product recommenders for online shops, playlist generators for video and audio services, or content recommenders for social networking platforms The main operationalization of this objective has been to concentrate on the capacity to numerically estimate customers' preferences for unseen objects. The purpose of recommenders is frequently stated as ''assist the customers identify relevant items.'' Which doctor to trust is one of the most frequently encountered worries among individuals when faced with any medical ailment. It is common knowledge that a person's health has a big impact on how happy they are. 58.99% of Americans have gone online for health-related information, according to a 2013 survey by the Pew Internet and American Life Project, with 35.6% of respondents concentrating on online medical condition diagnosis [2]. Every day, more individuals become concerned about issues related to health and medical diagnosis, but many people continue to perish as a result of medical mistakes. According to the administration's research, drug errors cause more than 200,000 fatalities annually in China and more than 100,000 in the USA. Doctors are at blame for more than 42% of drug errors because they write prescriptions based on their relatively limited experience. Finding qualified medical professionals to diagnose and treat medical disorders are therefore one of the most crucial choices a patient must make. The development of data mining and recommender technologies enables us to investigate possible knowledge from diagnosis history records, reviews, and ratings of medications in order to assist doctors in prescribing the right prescription and effectively reduce medication errors [3]. This data mining paper's goal is to develop and put into use a global disease prediction and drug recommendation system that integrates a variety of data mining technologies. We utilize a variety of prediction algorithms, together with NLP for sentiment analysis and recommendation, to merge data from diverse sources. The remainder of the report discusses data collection, pre-processing, methodology, findings, and finally the paper's conclusion and future work.

SCOPE OF THE PROJECT

The scope of this project encompasses the development and implementation of a data-driven system for personalized drug recommendations in the healthcare domain. The project will involve the utilization of data mining techniques to analyze large blocks of medical data, particularly focusing on symptom data extracted from hospital information systems (HIS) and patient reviews. The system will enable patients to input their symptoms, following which the data mining procedures will uncover significant patterns and trends within medical databases. The scope further includes the design and implementation of algorithms to process

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

patient reviews, ratings, and medical conditions, integrating probabilistic and weighted average methodologies to generate accurate drug recommendations. The project will entail comprehensive experimentation to validate the effectiveness and reliability of the proposed approach. Additionally, the scope extends to documenting the methodologies employed, including detailed descriptions of the models and strategies utilized, thereby providing a foundation for future studies and applications in the field of personalized medicine.

OBJECTIVE

The objective of this project is to develop a robust and efficient system for personalized drug recommendations in healthcare, leveraging data mining techniques to analyze symptom data and patient reviews. The primary aim is to empower patients by providing them with timely and accurate information about their health conditions and corresponding medications. Design and implement algorithms to extract relevant patterns and trends from large medical databases, focusing on symptom data obtained from hospital information systems (HIS). Develop a user-friendly interface that allows patients to input their symptoms and receive personalized drug recommendations based on the severity of symptoms, patient ratings, and medical conditions.

EXISTING SYSTEM:

The existing system aims to address health-related concerns by developing a global disease prediction and drug recommendation system. It utilizes data mining procedures to analyze and explore large blocks of medical data, with a focus on predicting diseases and recommending medications based on various algorithms and techniques.

Existing System Disadvantages:

- Addresses the challenge of medical mistakes and medication errors.
- Explores knowledge from diagnosis history records, reviews, and medication ratings

2. LITERATURE SURVEY

Title: Drug recommendation system for diabetes using a collaborative filtering and clustering approach: Development and performance evaluation,:

Author: L. F. G. Morales, P. Valdiviezo-Diaz, R. Reátegui, and L. Barba-Guaman.

Year: 2022.

Description: Diabetes is a public health problem worldwide. Although diabetes is a chronic and incurable disease, measures and treatments can be taken to control it and keep the patient stable. Diabetes has been the subject of extensive research, ranging from disease prevention to the use of technologies for its diagnosis and control. Health institutions obtain information required for the diagnosis of diabetes through various tests, and appropriate treatment is provided according to the diagnosis. These institutions have databases with large volumes of information that can be analyzed and used in different applications such as pattern discovery and outcome prediction, which can help health personnel in making decisions about treatments or determining the appropriate prescriptions for diabetes management. The aim of this study was to develop a drug recommendation system for patients with diabetes based on collaborative filtering and clustering techniques as a complement to the treatments given by the treating doctor. The data set used contains information from patients with diabetes available in the University of California Irvine Machine Learning Repository. Data mining techniques were applied for processing and analysis of the data set. Unsupervised learning techniques were used for dimensionality reduction and patient clustering. Drug predictions were obtained with a user-based collaborative filtering approach, which enabled creating a patient profile that can be compared with the profiles of other patients with similar characteristics.

Title: Automated disease diagnosis and precaution recommender system using supervised machine learning.

Author: F. Rustam, Z. Imtiaz, A. Mehmood, V. Rupapara, G. S. Choi, S. Din, and I. Ashraf, ‘

Year: 2022.

Description: Similar to many other professions, the medical field has undergone immense automation during the past decade. The complexity and rise of healthcare data led to a surge in artificial intelligence applications. Despite increased automation, such applications lack the desired accuracy and efficiency for healthcare problems. To address the aforementioned issue, this study presents an automatic health care system that can effectively substitute a doctor at an initial stage of diagnosis and help save time by recommending the necessary precautions. The proposed approach comprises two modules where Modul-1 aims at training the machine learning models using the disease symptoms dataset and their corresponding symptoms and precautions. Preprocessing and feature extraction are done as prerequisite steps. In Module-1 several algorithms are applied to the disease dataset such as support vector machine, random forest, extra trees classifier, logistic regression, multinomial naive Bayes, and decision tree. Module-2 interacts with the user (patient) through which the patient can describe the illness symptoms using a microphone. The voice data are transformed into text using the Google speech recognizer. The transformed data is later used with the trained model for disease prediction, as well as, recommending the precautions. The proposed approach achieves an accuracy of 99.9% during the real-time evaluation.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

Title: ‘‘A fair and safe usage drug recommendation system in medical emergencies by a stacked ANN.
Author : U. Bhimavarapu, N. Chintalapudi, and G. Battineni.
Year: 2022.
Description: The importance of online recommender systems for drugs, medical professionals, and hospitals is growing. Today, the majority of people use online consultations for drug recommendations for all types of health issues. Emergencies such as pandemics, floods, or cyclones can be helped by the medical recommender system. In the era of machine learning (ML), 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. Therefore, we present drug recommender systems with a stacked artificial neural network (ANN) model to improve the fairness and safety of treatment for infectious diseases. To reduce side effects, drugs are recommended based on a patient’s previous health profile, lifestyle, and habits. The proposed system produced results with 97.5% accuracy. A system such as this could be useful in recommending safe medicines to patients, especially during health emergencies.
Title: NLP based bi-directional recommendation system: Towards recommending jobs to job seekers and resumes to recruiters.
Author: S. A. Alsaif, M. S. Hidri, I. Ferjani, H. A. Eleraky, and A. Hidri.
Year: 2022.
Description: More than ten years ago, online job boards have provided their services to both job seekers and employers who want to hire potential candidates. The provided services are generally based on traditional information retrieval (IR) techniques which may not be appropriate for both job seekers and employers. The reason is that the number of produced results for job seekers may be enormous. Therefore, they are required to spend time reading and reviewing their finding criteria. Reciprocally, recruitment is a crucial process for every organization. Identifying potential candidates and matching them with job offers requires a wide range of expertise and knowledge. This article proposes a reciprocal recommendation based on bi-directional correspondence as a way to support both recruiters' and job seekers' work. Recruiters can find the best-fit candidates for every job position in their job postings, and job seekers can find the best-match jobs to match their resumes. We show how machine learning (ML) can solve problems in natural language processing (NLP) of text content and similarity scores depending on job offers in Saudi major cities scrapped from Indeed. For bi-directional matching, a similarity calculation based on the integration of explicit and implicit job information from two sides (recruiters and job seekers) has been used. The proposed system is evaluated using a resume/job offer dataset. The performance of generated recommendations is evaluated using ML decision support measures. Obtained results confirm that the proposed system can not only solve the problem of bi-directional recommendation but also improve the prediction accuracy.
Title: ‘A computer-based disease prediction and medicine recommendation system using machine learning approach.
Author: J. P. Gupta, A. Singh, and R. K. Kumar.
Year: 2021.
Description:
The combat between viral diseases and the human has been continuing since long history. But according to the evolution theory every entity of the world works for their survival always so, even the minute viruses also. Therefore, spreading infection through viruses are rapidly-evolving day by day and therefore it imposes a substantial burden on humans in terms of morbidity and mortality. Even though in this era, we have many advance techniques for diagnosis, prevention and treatment of infectious diseases, still the arrival of new diseases put a critical and emergence challenge to the global population. Recent example is the novel coronavirus, COVID-19, which was first found in Wuhan, China, and promptly became a global pandemic. There was no medicine available to cure the patient from this novel virus disease. In this tough situation, doctors and drug specialist were manually recommending the existing medicine based on the symptom occurs in the patients. During this process lots of infected people have died due to a lack of proper medicine. Therefore, in this work we have implemented a disease prediction system based on various symptoms of the disease. In addition to this, we came with an idea which can help the medicine industries towards the development of medicine for any viral disease using Machine Learning technique. Basically, this technique analyses the symptoms and predicts the best suitable medicine for any new disease. Moreover, this method, also predict the required composition of chemical elements that can be used by the medicine companies medicine to develop the new medicine under the supervisions of drug experts.

PROPOSED SYSTEM

The proposed system builds upon the existing one and aims to enhance the prediction accuracy and recommendation capabilities for diseases and medications. It incorporates a variety of data mining technologies and algorithms.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

Integration of NLP and Sentiment Analysis
Utilizes sentiment analysis based on NLP to analyze reviews.
Enhances the understanding of user sentiments and experiences related to medications.
Utilizes advanced prediction algorithms to enhance accuracy.

Data Collection and Pre-processing
Implements robust data collection and pre-processing techniques Implements robust data collection and pre-processing techniques
Incorporates methodologies to handle diverse data sources efficiently.

Proposed system advantages:

- Implements robust data collection and pre-processing techniques.
- Incorporates methodologies to handle diverse data sources efficiently

3. METHODOLOGY

This project focuses on the development of a data-driven system aimed at providing personalized drug recommendations within the healthcare domain. Leveraging advanced data mining techniques, the system analyzes extensive medical datasets, particularly focusing on symptom data extracted from hospital information systems (HIS) and patient reviews. By allowing patients to input their symptoms, the system employs sophisticated algorithms to uncover significant patterns and trends within the data, facilitating informed decision-making regarding medication choices. The system incorporates probabilistic and weighted average methodologies to process patient reviews and ratings, ensuring the accuracy and reliability of the drug recommendation process. Through comprehensive experimentation and validation, the project aims to demonstrate the efficacy and efficiency of the proposed approach in enhancing patient outcomes and optimizing medical treatment processes. Moreover, by documenting the methodologies employed, the project seeks to contribute to the advancement of personalized medicine and pave the way for future research and applications in the field.

1) DataSet:
A dataset is a structured collection of data, often presented in tabular or hierarchical format, used for analysis, research, or other purposes. It typically consists of rows representing individual observations or samples and columns representing variables or attributes. Datasets can vary widely in size and complexity, ranging from small, curated datasets used in academic research to massive, unstructured datasets collected from sources like social media or sensor networks for big data analysis.

2) Data Preprocessing:
Data preprocessing is a crucial step in data analysis and machine learning, involving cleaning, transforming, and organizing raw data into a format suitable for further analysis. This process typically includes handling missing values, removing duplicates, scaling features, encoding categorical variables, and splitting data into training and testing sets. By addressing inconsistencies and preparing the data effectively, preprocessing enhances the quality and reliability of downstream analyses and model predictions.

3) NLP , NLTK Technique:
Natural Language Processing (NLP) is a field of artificial intelligence that focuses on the interaction between computers and humans through natural language. NLTK (Natural Language Toolkit) is a widely-used Python library for NLP tasks, offering tools for tasks such as tokenization, stemming, lemmatization, part-of-speech tagging, and named entity recognition.

4) Data Parsing:
Data parsing is the process of extracting relevant information from raw data in a structured or unstructured format. It involves breaking down the data into smaller components, such as fields or tokens, and interpreting their meaning according to a predefined schema or set of rules.

5) Naïve Bayes Model:
The Naive Bayes model is a probabilistic machine learning algorithm based on Bayes' theorem and the assumption of conditional independence between features. It is commonly used for classification tasks, especially in natural language processing and text classification

6) Disease Prediction:
Disease prediction involves using various data-driven techniques, including machine learning and statistical analysis, to forecast the likelihood of an individual developing a particular illness or condition. This process typically utilizes medical data such as patient health records, genetic information, environmental factors, and lifestyle habits to build predictive models.

7) Drug Recommendation:
Drug recommendation systems utilize data-driven approaches, including machine learning and collaborative filtering techniques, to suggest suitable medications for patients based on their medical history, symptoms, genetic makeup, and other relevant factors. These systems analyze vast amounts of data, including drug efficacy, side effects, patient demographics, and treatment outcomes, to provide personalized and effective recommendations.

TECHNIQUE USED OR ALGORITHM USED
EXISTING TECHNIQUE: -
Broadly defined as a system anticipating customer ratings for specific medications.
Provides rankings based on predicted customer preferences.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

Integrates probabilistic and weighted average methodologies for drug recommendations

PROPOSED TECHNIQUE USED OR

- Enhances the understanding of user sentiments and experiences related to medications along with the Machine Learning algorithms and NLP.
- Deploys four distinct prototypes for disease prediction.
- Utilizes advanced prediction algorithms to enhance accuracy.
- Aims to create a comprehensive global disease prediction system.
- Integrates diverse data sources for a more holistic prediction approach.

ALGORITHM USED:

.SYSTEM ARCHITECTURE:

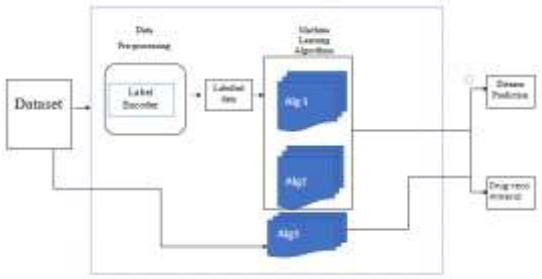

Fig 4.11: System Architecture

FUTURE ENHANCEMENTS:

Key future scope can definitely be improving the accuracies of the prediction and recommender model using deep neural networks by using larger data. Furthermore, extending the scope of the system to encompass a broader range of medical conditions and medications could enhance its utility and impact in clinical settings. Collaboration with healthcare institutions and pharmaceutical companies may facilitate the integration of the system into existing healthcare infrastructure, enabling seamless adoption and widespread use.

CONCLUSION

Drug recommendation systems are a common technology in today's online services, and as demand for these services grows, there is an increasing need to automate the processes. As a result, we have created a medication recommendation system. The main conclusions from our project are listed below. Successfully created a drug recommendation prototype that prescribes medicines with potential adverse effects based on user-inputted symptoms. For the execution of this project, we created three models. a model for sentiment analysis, one for predicting diseases, and one for making recommendations. Tested several strategies for each of the three models. Each of the three models provided accurate results, adding to the drug recommendation model's overall dependability.

REFERENCES:

[1] F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh, ''Recommendation systems: Principles, methods and evaluation,'' Egyptian Informat. J., vol. 16, no. 3, pp. 261–273, Nov. 2015.

[2] M. A. N. Banu and B. Gomathy, ''Disease predicting system using data mining techniques,'' Int. J. Tech. Res. Appl., vol. 1, no. 5, pp. 41–45, 2013.

[3] H. Wang, Q. Gu, J. Wei, Z. Cao, and Q. Liu, ''Mining drug-disease relationships as a complement to medical genetics-based drug repositioning: Where a recommendation system meets genome-wide association studies,'' Clin. Pharmacol. Therapeutics, vol. 97, no. 5, pp. 451–454, May 2015.

[4] S. A. Alsaif, M. S. Hidri, I. Ferjani, H. A. Eleraky, and A. Hidri, ''NLPbased bi-directional recommendation system: Towards recommending jobs to job seekers and resumes to recruiters,'' Big Data Cognit. Comput., vol. 6, no. 4, p. 147, Dec. 2022.

[5] J. P. Gupta, A. Singh, and R. K. Kumar, ''A computer-based disease prediction and medicine recommendation system using machine learning approach,'' Int. J. Adv. Res. Eng. Technol. (IJARET), vol. 12, no. 3, pp. 673–683, 2021.

[6] Y. Bao and X. Jiang, ''An intelligent medicine recommender system framework,'' in Proc. IEEE 11th Conf. Ind. Electron. Appl. (ICIEA), Jun. 2016, pp. 1383–1388.

[7] Q. Zhang, G. Zhang, J. Lu, and D. Wu, ''A framework of hybrid recommender system for personalized clinical prescription,'' in Proc. 10th Int. Conf. Intell. Syst. Knowl. Eng. (ISKE), Nov. 2015, pp. 189–195.

[8] U. Bhimavarapu, N. Chintalapudi, and G. Battineni, ''A fair and safe usage drug recommendation system in medical emergencies by a stacked ANN,'' Algorithms, vol. 15, no. 6, p. 186, May 2022.

[9] J. Chen, K. Li, H. Rong, K. Bilal, N. Yang, and K. Li, ''A disease diagnosis and treatment recommendation system based on big data mining and cloud computing,'' Inf. Sci., vol. 435, pp. 124–149, Apr. 2018.

[10] I. Kononenko, I. Bratko, and M. Kukar, ''Application of machine learning to medical diagnosis,'' Mach. Learn. Data Mining, Methods Appl., vol. 389, p. 408, Jun. 1997.

[11] C. R. Olsen, R. J. Mentz, K. J. Anstrom, D. Page, and P. A. Patel, ''Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure,'' Amer. Heart J., vol. 229, pp. 1–17, Nov. 2020.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

[12] A. S. Hussein, W. M. Omar, X. Li, and M. Ati, ''Efficient chronic disease diagnosis prediction and recommendation system,'' in Proc. IEEE-EMBS Conf. Biomed. Eng. Sci., Dec. 2012, pp. 209–214.
[13] F. Rustam, Z. Imtiaz, A. Mehmood, V. Rupapara, G. S. Choi, S. Din, and I. Ashraf, ''Automated disease diagnosis and precaution recommender system using supervised machine learning,'' Multimedia Tools Appl., vol. 81, no. 22, pp. 31929–31952, Sep. 2022.
[14] S. Bhat and K. Aishwarya, ''Item-based hybrid recommender system for newly marketed pharmaceutical drugs,'' in Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI), Aug. 2013, pp. 2107–2111.
[15] K. Feldman, D. Davis, and N. V. Chawla, ''Scaling and contextualizing personalized healthcare: A case study of disease prediction algorithm integration,'' J. Biomed. Informat., vol. 57, pp. 377–385, Oct. 2015. [16] Y. Zhang, S. Fong, J. Fiaidhi, and S. Mohammed, ''Real-time clinical decision support system with data stream mining,'' J. Biomed. Biotechnol., vol. 2012, pp. 1–8, May 2012.
[17] P. C. Austin, J. V. Tu, J. E. Ho, D. Levy, and D. S. Lee, ''Using methods from the data-mining and machine-learning literature for disease classification and prediction: A case study examining classification of heart failure subtypes,'' J. Clin. Epidemiol., vol. 66, no. 4, pp. 398–407, Apr. 2013.
[18] E. AbuKhousa and P. Campbell, ''Predictive data mining to support clinical decisions: An overview of heart disease prediction systems,'' in Proc. Int. Conf. Innov. Inf. Technol. (IIT), Mar. 2012, pp. 267–272. [19] T. N. T. Tran, A. Felfernig, C. Trattner, and A. Holzinger, ''Recommender systems in the healthcare domain: State-of-the-art and research issues,'' J. Intell. Inf. Syst., vol. 57, no. 1, pp. 171–201, Aug. 2021. [20] L. F. G. Morales, P. Valdiviezo-Diaz, R. Reátegui, and L. Barba-Guaman, ''Drug recommendation system for diabetes using a collaborative filtering and clustering approach: Development and performance evaluation,'' J. Med. Internet Res., vol. 24, no. 7, Jul. 2022, Art. no. e37233.
[21] Y. Zhang, M. Chen, D. Huang, D. Wu, and Y. Li, ''iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization,'' Future Gener. Comput. Syst., vol. 66, pp. 30–35, Jan. 2017.
[22] M. Kuanr, P. Mohapatra, and J. Piri, ''Health recommender system for cervical cancer prognosis in women,'' in Proc. 6th Int. Conf. Inventive Comput. Technol. (ICICT), Jan. 2021, pp. 673–679.
[23] Q. Han, M. Ji, I. M. de Rituerto de Troya, M. Gaur, and L. Zejnilovic, ''A hybrid recommender system for patient-doctor matchmaking in primary care,'' in Proc. IEEE 5th Int. Conf. Data Sci. Adv. Anal. (DSAA), Oct. 2018, pp. 481–490.
[24] V. Mudaliar, P. Savaridaasan, and S. Garg, ''Disease prediction and drug recommendation Android application using data mining (virtual doctor),'' Int. J. Recent Technol. Eng. (IJRTE), vol. 8, no. 3, pp. 6996–7001, Sep. 2019.