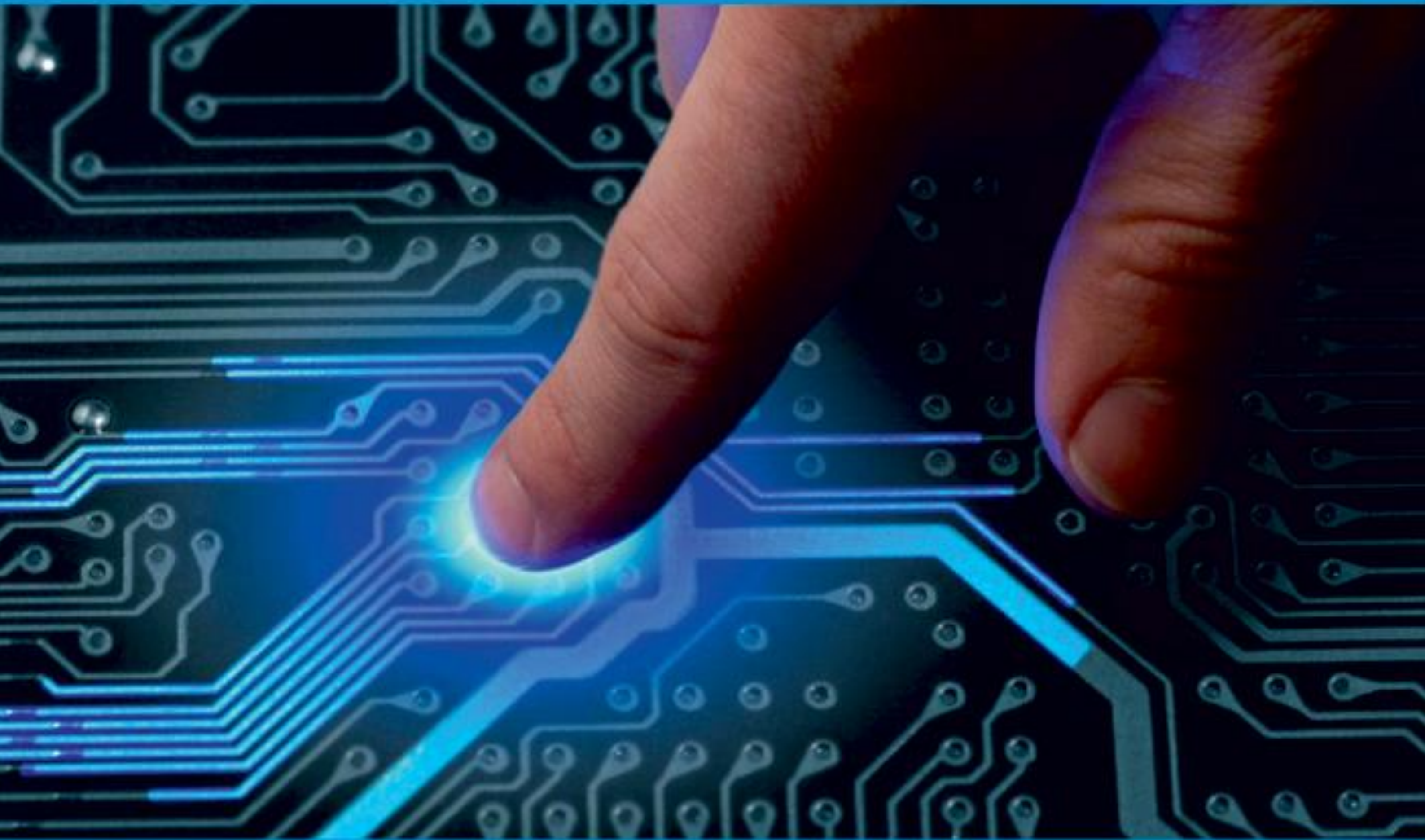
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205148 |

Drug Recommendation and Disease Prediction System Using Machine Learning Approaches

B. Manasa*[1], B. Devi Priya*[2], E. Suneetha*[3], K. Sai Pragna*[4], K. Penchalaiah*[5]

B. Tech Students, Department of Computer Science Engineering, NBKRIST, Vidyanagar, India *1,2,3,4

Assistant professor, Department of Computer Science Engineering, NBKRIST, Vidyanagar, India *5,

ABSTRACT: Online recommender systems are being used increasingly often for hospitals, medical professionals, and drugs. Today, the great majority of consumers look online before asking their doctors for prescription suggestions for a range of health conditions. The medical suggestion system can be valuable when pandemics, floods, or cyclones hit. In the age of Machine Learning (ML), recommender systems give more accurate, precise, and reliable clinical predictions while using less resources. The medicine recommendation system gives the patient reliable information about the medication, the dosage, and any possible adverse effects. Medication is given based on the patient's symptoms, blood pressure, diabetes, temperature, and other parameters. Drug recommendation systems provide precise information at any time while improving the performance, integrity, and privacy of patient data in the decision-making process. Recommender system, the decision tree produces the most accurate results. In times of medical emergency, a drug recommendation system is helpful for giving patients recommendations for safe medications.

KEYWORDS: Data Mining, Drug Recommendation System, Random Forest, SVC.

I.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 wellknown 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 thisobjective 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. 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 error. 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.

II.LITERATURE SURVEY

- An item-based hybrid recommender system was created by Bhat and Aishwarya [1] to increase the precision of drug recommendations for recently released pharmaceuticals. To make precise medicine recommendations, the system integrates content-based and collaborative filtering algorithms. The proposed model's accuracy is 75%.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205148 |

• Incorporating disease prediction algorithms into customised healthcare can help it be scaled and contextualised, as Feldman et al. [2] demonstrate. They also emphasised the significance of scalability, pointing out that for customised healthcare to be successfully integrated, it is important to take into account the influence of data size and the technical infrastructure required to support it.

• A new system that can assess medical data streams and produce real-time predictions in clinical decision support was introduced by Zhang et al. [3]. This system is based on the VFDT (Very Fast Decision Tree) stream mining technique. The authors also go into some of the difficulties that come with data stream mining, like data size, data complexity, and data accuracy. • In a case study, Austin et al. [4] used several data mining and machine learning techniques to categorise and forecast the disease of heart failure. They compared alternative flexible classification schemes, such as bootstrap aggregation (bagging), boosting, and random forests, to the conventional classification and regression trees and discovered that the flexible tree-based techniques from the data-mining literature offer a significant improvement in prediction and classification of heart failure subtype.

• In order to support clinical judgements in the field of cardiac disease prediction, AbuKhousa et al. [5] assessed the state of research and development in predictive data mining. The potential ramifications of five alternative models of data mining techniques have been examined. Poor generalization capacity is a significant unresolved problem for data mining in the healthcare sector due to the dearth of input data and the high cost of re-processing.

• An overview of the present state of research on recommender systems in the healthcare industry is given by Tran et al. [6]. They talk on how existing Systems in the healthcare industry are structured and designed, as well as how this affects patient care. This in depth analysis offers insights into scenario-based recommendations, approach based recommendations, and a variety of algorithm-based recommendations.

• The review contains recommendations for foods, medications, healthcare professionals, healthcare services, and health status forecasts. Morales et al. [7] designed a drug recommendation system based on collaborative filtering and clustering methods as an addition to the medications provided by the prescribing physician for diabetic patients. To conduct the studies, a set of diabetes patient data is gathered from the University of California Irvine Machine Learning Repository. For dimensionality reduction, principal component analysis is employed as a method, and user-based collaborative filtering is used for medication prediction. The proposed recommendation system produces a result that is satisfactory, with an accuracy of 0.61 and a mean squared error metric of 0.51. • A innovative healthcare recommendation system named iDoctor was devised by Zhang et al. [8] and is based on hybrid matrix factorization techniques. Sentiment analysis is employed to eliminate the emotional component of user reviews and to update the actual user ratings. Latent Dirichlet Allocation is used to extract user preferences and doctor features, which are then added to a traditional matrix factorization. The suggested model's accuracy is higher when compared to the current healthcare recommendation system using open real datasets found on the crowd-sourced review website Yelp. Four techniques—Hybrid Matrix Factorization (HMF), Basic Matrix Factorization (BMF), and Item-based and User based Collaborative Filtering—are compared. According to the outcome, the RSME (Root Mean Square Error) HMF is the lowest of all.

• The application of a health recommender system to increase the precision and effectiveness of women's cervical cancer prognosis was studied by Kuanr et al. [9] in 2021. LR (Logistic Regression), SVC (Support Vector Classifier),
DT (Decision Tree), KNN (KNearest Neighbor), GNB (Gaussian Naive Bayes), XG Boost (eXtreme Gradient Boosting), and GBM (Gradient Boosting Machine) are the seven classifiers utilised for the model construction.

• The findings also suggest that models with GBM classifiers perform admirably and the model with Decision Tree classifiers exhibit the highest accuracy. Han et al. [10] suggested a hybrid recommender system that combines contentbased and collaborative filtering strategies for patient-doctor matchmaking in primary care. The results demonstrate that, when compared to both the heuristic baseline (which provides 37% •
accuracy) and a traditional collaborative filtering (which provides 69% accuracy) recommender system, the hybrid model provides greater predicted accuracy of 80%. The model was evaluated on a dataset of large number of doctor-patient interactions.

III.MOTIVATION

A drug recommendation system can motivate healthcare providers and patients by improving treatment outcomes, reducing adverse effects, enhancing patient compliance, and ultimately, saving lives. It streamlines the process of selecting the most suitable medication tailored to individual needs, increasing efficiency in healthcare delivery.
The exponential development of the web-based industry made the item reviews as the integral factor in purchasing the items. Reviews and ratings of an item plays a major role in getting choose by or recommend to the individuals. Similarly in the healthcare domain, Drug reviews play a very important role in providing crucial medical care

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205148 |

information for both healthcare professionals and consumers. Analyzing the drug reviews will not only be useful to recommend the best drugs for the patients but also helps the pharmacy companies and healthcare professionals to study and monitor the adverse side effects post marketing the drug thereby to improve the consumer safety.

IV.OBJECTIVE

Predicting Disease Risk: Develop algorithms to analyse diverse patient data, such as medical records, genetics, lifestyle factors, and environmental influences. These algorithms will predict the likelihood of developing specific diseases or conditions before their onset.
Recommending Optimal Treatments: Create a robust drug recommendation system that considers individual patient characteristics and predicted disease risks. This system will suggest personalized treatment plans, including medication options and lifestyle modifications, to optimize therapeutic outcomes.
Improving Patient Care: Enhance the quality of patient care by providing tailored interventions that address individual health needs and risk factors. By leveraging predictive analytics, healthcare providers can intervene early, potentially preventing the progression of diseases and improving overall patient well-being.

V. DATASET AND PRE-PROCESSING

HIPPA provides protection for healthcare information. It is against the law to disclose a patient's medical records without that patient's consent. Government health records and databases required a number of licenses to access. As a result, we are employing datasets for our research that were easily accessible online and ready for download.

A. DATASET ONE

1) DATA GATHERING

Medical records 10 yrs. - dataset by arvin6 | data.world is where the 10-year medical record dataset was found.Four CSV files make up the package:

1. encounter.csv
2. encounter_dx.csv
3. lab_results.csv
4. medication_fulfillment.csv

There are 17 columns and 1176 rows in the encounter.csv file, encounter dx. Lab results.csv has 7509 rows and 21 columns, medical fulfillment.csv has 5447 rows and 28 columns, and csv has 3063 rows and 6 columns. We pre-processed and merged the dataset to meet our needs in order to gain insightful information and determine whether the dataset contains the necessary data to address the problem statement.

2) DATA PREPROCESSING

We pre-processed each table individually and then joined the tables to create a single merged dataset with the necessary columns after acquiring the raw data and knowing the structure of all four CSV files. In order to pre-process the data, we ran specific instructions. For example, the number of unique values, the count of each row, and the columns that were removed or combined to make the data more useful because they were unnecessary or did not include any data.
After performing the necessary pre-processing, we discovered that the four tables were controlled using the Star Schema Model, with the fact table being medical fulfillment.csv and the dimension tables being ecounter.csv, encounter dx.csv, and lab results.csv. The one-to-many relationship is followed by the star schema. The four tables were combined into one table once we identified the Primary Key and Foreign Key. The primary key in the medication fulfillment table is
'Ecounter ID'. We then used a left join on "Encounter ID" in a SQL query to combine the Medication fulfillment table with the encounter dx table's severity and description columns. There are 1176 rows in the ensuing table.
Then, we performed a query to verify the count of each row in the "order ID" column to see if it was the primary key for lab results.csv. We discovered that it was not since the number of unique values did not correspond to the number of rows in the table. We then discovered that the composite primary key consists of two columns "Order ID" and "Result LOINC." We did not use any columns for the combined dataset because none of the lab results.csv's columns were helpful.
We determined that the primary key for the remaining encounter.csv file is "Encounter ID," therefore we extracted the CC column and combined it using a left join on
"Encounter ID." We now have a fully combined dataset with all of the necessary columns. There are 1176 rows in it. We took the Drug Name, description, severity, and CC from the combined dataset and grouped them to determine the

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205148 |

total number of each drug and the number of linked diseases and descriptions. As seen in the Figure 1 below, the extracted columns contain a significant number of null values.

	Drug_Name	description	severity	CC	cnt
0	[illegible]	Chronic Obstructive Pulmonary Disease	critical	critical shortness of breath	[illegible]
1	Ciprofloxacin	None	None	None	[illegible]
2	Isotonic Saline (0.9%)	None	None	None	101
3	Lisinopril	None	None	None	[illegible]
4	Potassium Cl	Type 1 Diabetes	severe	severe increased thirst	78
...	...	...	...	...	...
76	metoprolol	Hypertension	severe	moderate difficulty walking	1
77	oxycodone-acetaminophen 10-325	None	None	None	1
78	trimethoprim	None	None	None	1
79	trimethoprim	Pyelonephritis	severe	Pyelonephritis	1
80	valsartan	Chronic Congestive Heart Failure	severe	mild palpitations	1

81 rows × 5 columns

FIGURE 1. Drug name grouped by description, severity, and CC.

In order to get the total number of rows with null values associated with a medicine name, we conducted another query. We were only left with 416 rows of pertinent data to train our classification model because a total of 764 rows had null values. We therefore came to the conclusion that this dataset does not fulfill the requirement and searched for additional datasets that would be consistent with the solution of our issue statement.

B. DATASET TWO

1) DATA GATHERING

We forecast diseases based on symptoms in order to make reliable drug recommendations, which are subsequently based on ratings. We have made an effort to assemble data from two major datasets for this purpose.

a: SYMPTOMS DATASET

The Disease-Symptom Knowledge Database, which is a knowledge database of disease-symptom associations generated by an automated method based on information in textual discharge summaries of patients at New York-Presbyterian Hospital admitted during 2004, provided the dataset. This dataset has three columns as shown in Figure 2:

1. Disease
2. Count of Disease Occurrence
3. Symptoms

	Disease	Count of Disease Occurrence	Sympto
0	UMLS:C0020538_hypertensive disease	3363.0	UMLS:C0008031_pain che
1	NaN	NaN	UMLS:C0392680_shortness brea
2	NaN	NaN	UMLS:C0012833_dizzine
3	NaN	NaN	UMLS:C0004093_asthen
4	NaN	NaN	UMLS:C0085639_f

FIGURE 2. Raw symptoms dataset.

This dataset contains 405 symptoms and 149 distinct diseases. Each disease has 4-5 symptoms that go along with it. In order to train models to categorize and forecast the disease, this dataset is provided for preprocessing.

b: DRUG REVIEW DATASET

Using the projected disease as an input, this dataset is utilized to suggest suitable medications based on reviews and ratings (Sentiment Analysis). The UCI Machine Learning Repository for Drug Review, which offers patient reviews on particular medications together with information on linked ailments and a 10star patient rating indicating overall patient happiness, is where the data set is acquired. Due to the fact that the two datasets in the repository (Test and Train) had the same amount of columns, they were combined for analysis and visualization. It has 213869 rows and 7 columns: ID, drug name, condition, Review, Rating, Date and Useful count as in Figure 3.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205148 |

FIGURE 3. Raw drug review dataset.

This dataset includes 916 unique Conditions (Diseases), 3671 unique Drug names, and ratings and reviews that match to the medicine names. To obtain more information for efficient medicine recommendations, this dataset is then pre-processed and displayed.

c: SIDE EFFECTS DATASET

We were able to successfully include this dataset of adverse effects for particular medications in order to assist patients in understanding the risks associated with the medication that is being suggested. Once more, raw data from druglib.com and the UCI Machine Learning Archive for Adverse Effects of Medications were combined to create this dataset. Because of this, just the "Side Effects" column from this dataset will be integrated with the side effects from the other two datasets and druglib.com as shown in Figure 4.

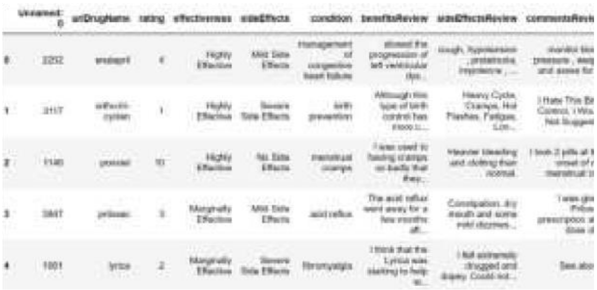


FIGURE 4. Raw dataset containing Side effects of drugs.

2) DATA PREPROCESSING

We used several datasets and worked with them in this paper. The datasets were all acquired in their raw form. A few standard procedures and tests were run on all of the datasets to pre-process them. They were:

1. The number of null and missing values in each dataset was initially counted.
2. Every one of these values was either dealt with or removed from the dataset.
3. The unique values and frequencies for each column were then determined.
4. The dataset was displayed using standard libraries, and any outliers were discovered.

The dataset was cleaned of any extraneous data.

a: SYMPTOMS DATASET

This dataset needed to be cleansed in order to yield useful information. The "Count" column was first removed since the data it contained was not pertinent to this paper. The drop function was then used to handle the null entries in the Disease column. The disease and symptom columns were cleaned to remove extraneous data and leave only the name. The processed data is shown in Figure 5.

We have changed it into a new CSV format file with symptoms as the columns and diseases as the rows in order to categorize the symptoms according to the diseases. We mapped every symptom to every disease using one hot

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205148 |

	Disease	Symptom
0	hypertensive disease	[pain chest, shortness of breath, dizziness, a...
1	diabetes	[polyuria, polydypsia, shortness of breath, pa...
2	depression mental	[feeling suicidal, suicidal, hallucinations au...
3	depressive disorder	[feeling suicidal, suicidal, hallucinations au...
4	coronary arteriosclerosis	[pain chest, angina pectoris, shortness of bre...

FIGURE 5. Symptoms dataset after pre-processing.

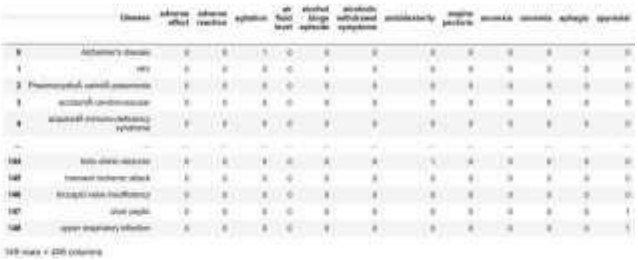


FIGURE 6. Dataset after marking symptoms present for a disease as 1 else 0.

encoding, adding value 1 if the disease was present and 0 otherwise. The one-hot encoded dataset is shown in the Figure 6 below. When symptoms are provided as input, this will assist us in predicting the diseases.

b: DRUG REVIEW DATASET

In order to visualize and analyze the data on a larger dataset, the two sets in this dataset—Train and Test— were combined. Also, they could be joined easily because they both had the same columns. The resulting dataset was fairly clean and did not need much preprocessing. However, a few rows with null values were removed, and new names were given to the columns. The displaying challenge was intriguing because the dataset has a lot of information. The results of the pharmaceuticals with the most evaluations, the most wellliked drugs, the most prevalent ailments, etc. were plotted on numerous different graphs. Figure 7 below illustrates one such representation, which includes the names of a few of the most often used drugs:

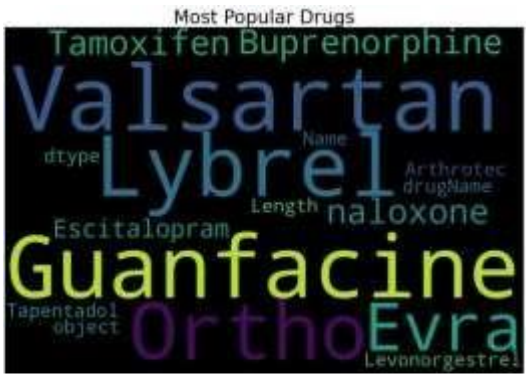


FIGURE 7. Visualizing the most popular drugs based on the ratings.

c: SIDE EFFECTS DATASET

This dataset was clean and has a lot of data that is similar to that in the Drug Review dataset. The handling of a few null values and the removal of unnecessary columns. By combining this information with the drug review dataset, it is possible to map solely the adverse effects of the particular medications.

d: MERGED DATASET

For the purpose of making a final medicine prediction, the data set containing the symptoms and the reviews are combined as in Figure 8.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205148 |

	Drug	Disease	Review	Rating	UsefulCount	Symptoms
0	Aspirin	transient ischemic attack	"No side effects, easy to take, no more sympt...	10	10	['speech slurred', 'dysarthria', 'facial pares...
1	Clopidogrel	transient ischemic attack	"I've been taking this medicine for a lit...	10	8	['speech slurred', 'dysarthria', 'facial pares...
2	Clopidogrel	transient ischemic attack	"I took ibuprofen (2 caps at night for severe ...	6	13	['speech slurred', 'dysarthria', 'facial pares...
3	Clopidogrel	transient ischemic attack	"After my VAD Stroke I am on plavix. I have a...	5	9	['speech slurred', 'dysarthria', 'facial pares...
4	Bayer Children's Aspirin	transient ischemic attack	"No side effects, easy to take, no more sympt...	10	10	['speech slurred', 'dysarthria', 'facial pares...

FIGURE 8. Merged dataset.

VI. SYSTEM ARCHITECTURE

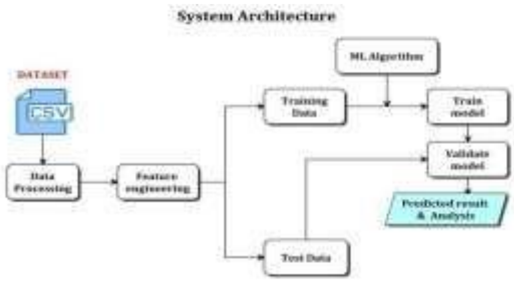


FIGURE 9. System Architecture

This architecture provides a foundation for building an intelligent system that can assist healthcare providers in making informed decisions about disease diagnosis and treatment selection based on individual patient characteristics and medical history.

VII. EXPERIMENTAL RESULTS

A. DISEASE PREDICTION

We have also utilized this method to make predictions for all 4 models, and the accuracy obtained is as follows: Decision Tree: 88.24%, Multinomial NB:
87.98%, Random Forest: 81.88%, Gaussian NB: 88.34%, and SVM: 88.61%. Each classifier has done well, and the condition that was predicted using three symptoms as input features is likewise accurate. When we used symptoms from various diseases to try and predict the disease, the model was 9 out of 10 times accurate. We utilized this strategy for our final disease prediction because the outcomes were pretty excellent and the calculation time was also reduced. The prediction result is shown in Figure 10.

Predicted disease with all models

```
dt.predict(arr)
array(['delirium'], dtype=object)

mnb.predict(arr)
array(['parkinson disease'], dtype='<U36')

rfc.predict(arr)
array(['HIV'], dtype=object)

naivebys.predict(arr)
array(['parkinson disease'], dtype='<U36')

svm.predict(arr)
array(['parkinson disease'], dtype=object)
```

FIGURE 10. Disease prediction using an approach (Prediction using all symptom).

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205148 |

B. DRUG RECOMMENDATION

We have tried with a method for the drug recommendation, the final suggestion was made by combining the usable count factor from the probabilistic technique with the weighted average approach. Many factors were taken into account before suggesting the medication:

Several medications are available for the same condition. Hence, we were able to filter out the negative and neutral evaluations using sentiment analysis, leaving us with only favorable ones.

This method produced fairly excellent findings and evaluations. Nevertheless, there isn't a concrete dataset to compare

The recommended Drugs for the given Disease is:

	Drug	Disease	Prob. of Side Effect	Side Effects
1	PROLIA	osteoporosis	0.1	Back pain, blistering, crusting, cracked, dry ...
8	DENOSUMAB	osteoporosis	0.4	Back pain, blistering, crusting, cracked, dry ...
2	TERIPARATIDE	osteoporosis	0.8	abdominal pain, confusion, constipation, depre...
3	FORTEO	osteoporosis	0.8	abdominal pain, confusion, constipation, depre...
4	ZOLEDRONIC ACID	osteoporosis	0.8	Agitation, blurred vision, cough, depression, ...

FIGURE 11. List of recommended drugs along with possible side effects for osteoporosis ranked based on probabilistic score.

the suggested outcomes as this list of medications for osteoporosis is advised together with any potential adverse effects, sorted by probabilistic score. Figure 11 shows the list of recommended drugs along with possible side effects for osteoporosis ranked based on probabilistic score.

The outcomes of this strategy can be contrasted with a list of osteoporosis medications that are typically prescribed.

VIII. DISCUSSION

In a single file, we have bundled all the functions used in this paper with the most precise techniques. This programme forecasts the disease using symptoms as input. This is then used as an input by the drug recommender, which subsequently gives the prescribed medication as well as a list of its side effects as an output.

IX. CONCLUSION AND FUTURE SCOPE

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.

1. Successfully created a drug recommendation prototypethat prescribes medicines with potential adverse effects based on user-inputted symptoms.
2. For the execution of this project, we created three models.a model for sentiment analysis, one for predicting diseases, and one for making recommendations.
3. Tested several strategies for each of the three models.
4. Each of the three models provided accurate results,adding to the drug recommendation model's overall dependability.

One key future scope can definitely be improving the accuracies of the prediction and recommender model using deep neural networks by using larger data.

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, "NLPbasedbi-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.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2024.1205148 |

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. 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,
12. ‘‘Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure,’’ *Amer. Heart J.*, vol. 229, pp. 1–17, Nov. 2020.
13. S. Hussein, W. M. Omar, X. Li, and M. Ati, ‘‘Efficient chronic disease diagnosis prediction and recommendation system,’’ in *Proc. IEEEEMBS Conf. Biomed. Eng. Sci.*, Dec. 2012, pp. 209–214.
14. 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.
15. 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.
16. 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.