**1.INTRODUCTION**

Humans use some form of language in order to communicate verbally or through written text. Computers need to comprehend human-used natural languages in order to engage with humans. Making computers capable of learning, processing, and manipulating natural languages is the goal of natural language processing.

Computers and basic human languages interact in the interdisciplinary approach of natural language processing (NLP). NLP-powered software helps us in our daily lives in various ways. One of the important applications is Drug recommendation system. There are several drugs available for a disease, and consumers usually struggle to identify medications for their diseases.

So, this Drug recommendation is helpful for patients by predicting patient’s outcome using drug reviews. The three primary data filtering techniques used by recommender systems—content-based filtering, collaborative filtering, and hybrid approaches—are used to handle the problem of large volume of data and to recommend to users the elements from dynamically created data that they will find interesting.

Healthcare prediction and recommender systems are intended to make a precise prediction of the disease and offer relevant precautions to the patients. With a large amount of daily generated data and information surcharge, healthcare recommender systems gained a particularly important role as they can enhance the prediction accuracy and suggest relevant precautions for the patients at the same time.

**2.LITERATURE SURVEY**

Following research papers are studied in details to understands the proposed recommendation technique and experimental result for predicting the output.

**2.1. Rustam, F., Imtiaz, Z., Mehmood, A. *et al.* Automated disease diagnosis and precaution recommender system using supervised machine learning. MultimedToolsAppl (2022).**

In this paper, we created a framework for recommending precautions and disease indicators using data mining techniques. Association rules, classification, clustering, prediction, and sequential models are some of the most popular data mining techniques that can be vital in health recommender systems.

For prediction, it is suggested to apply a machine learning approach in which the user communicates to the system, which is later translated to textual input and pre-processed.

A big-data technique is adopted for healthcare recommendation which compares patients generated data with the data cleansed by the proposed system. The data cleansing is performed using Apache Spark and IBM Watson services. This method auto-upgrades itself with the inflow of data and is a self-learning method. It aims at providing an interactive interface for the patients to monitor their symptoms, and follow recommendations generated from the machine learning model.

Support vector machine (SVM), random forest (RF), extra trees classifier (ETC), logistic regression (LR), multinomial naive Bayes (MNB), and decision tree are some of the machine learning approaches that have had their accuracy rate examined (DT).

**2.2. Bhimavarapu, U,; Chintalapudi, N.; Battineni, G. A Fair and Safe Usage Drug Recommendation System in Medical Emergencies by a Stacked ANN. Algorithms 2022,15,186.**

In this study, we aimed to create a drug recommender system (DRS) for various diseases in order to preserve the health and longevity of patients. By increasing the recommendation accuracy and incorporating ML knowledge, we addressed the unfairness in the use of drugs by DRS for severe chronic conditions.

Various criteria, including accuracy, sensitivity, and specificity, were used to determine the system performance. To confirm its effectiveness, the performance was also compared to that of other ML models that were already in use. A new model learns how to combine the predictions from several different current models in the most effective way through stacked generalization, an ensemble method.

In this work, we proposed the development of a drug recommender system (DRS) for different diseases to maintain good patient health and longevity. We addressed the unfairness in drug usage by DRS for severe chronic diseases by improving the recommendation accuracy by the integration of ML knowledge.

The drug recommender system can prescribe medication in accordance with a particular condition by integrating sentiment analysis and feature engineering.

**2.3. Sahoo, Abhaya & Pradhan, Chittaranjan & Barik, Dr. Rabindra & Dubey, Harishchandra. (2019). DeepReco: Deep Learning Based Health Recommender System Using Collaborative Filtering. Computation. 7. 25. 10.3390/computation7020025.**

The collaborative filtering-based health recommender system, that gives useful information to patients based on an item’s profile, is described in this work. Patients express their preferences for particular items, which are reflected as a utility matrix.

This matrix provides the value of each patient-item pair, which indicates the level of that patient’s preferences for those particular items.

By constructing a patient-item matrix of patient preferences for items, collaborative filtering predicts unknown possibilities. The patient-item matrix is compared to the patients’ preferences and interests to see how similar the profiles of different patients are. The Collaborative filtering can be utilised for either prediction or recommendation in the recommendation systems.

Using this recommender system process, we can increase our sales productivity in the market. While the preferences made by customers can be described as being low-risk, choices made in other sectors may have more intense ramifications for the end patient. In particular, in the sector of healthcare, choices can be life-threatening as they are concerned with the life and safety of patients.

**2.4. Granda Morales L, Valdiviezo-Diaz P, Reátegui R, Barba-Guaman Drug Recommendation System for Diabetes Using a Collaborative Filtering and Clustering Approach: Development and Performance EvaluationJ Med Internet Res 2022;24(7): e37233**

The aim of this study was to complement previous diabetes-related studies by first analysing data related to patients with diabetes to obtain important information for the management of this disease, followed by identifying groups of patients who share similar characteristics, which could enable discovering patterns of interest that can support decision-making.im of this study was to complement previous diabetes-related studies by first analysing data related to patients with diabetes to obtain important information for the management of this disease, followed by identifying groups of patients who share similar characteristics, which could enable discovering patterns of interest that can support decision-making.

As an addition to the medications recommended by the doctor, we propose a medicine recommendation system in this research for diabetic patients based on collaborative filtering and clustering approaches. Here, the data set was processed and analysed using data mining techniques, and dimensionality reduction and patient clustering were achieved using unsupervised learning techniques.

Collaborative filtering approach was used to get drug predictions, allowing the creation of patient profiles that could be compared to those of other patients who shared the same qualities. Finally, the performance of the system was evaluated using metrics.

**2.5. Ye, Q., Hsieh, CY., Yang, Z. et al*.* A unified drug–target interaction prediction framework based on knowledge graph and recommendation system. NatCommun 12, 6775 (2021).**

In this paper, we develop KGE\_NFM, a unified framework for DTI, which refers to the recognition of interactions between chemical compounds and the protein targets in the human body.

In addition, network-based methods have been developed by incorporating multiple data sources, such as drug–target interactions, drug–drug interactions, and protein–protein interactions, into one framework for DTI prediction. In these networks, nodes can be drugs or proteins and edges are the indicators for the interactions or similarities between the connected nodes

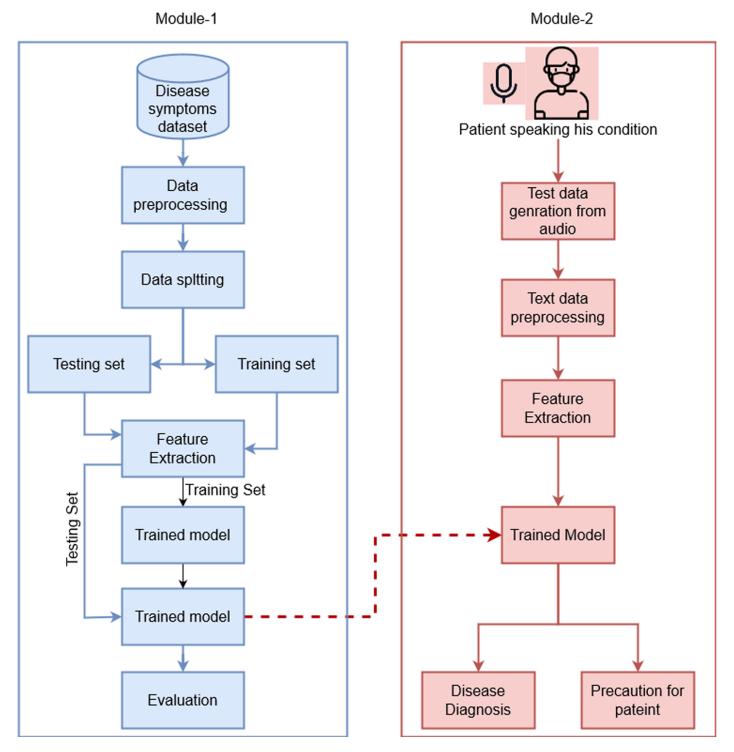
Initially this framework acquires a low-dimensional representation for each item in the KG, and then use a neural factorization machine to combine the multimodal data. Structure-based techniques, ligand-based approaches, and hybrid approaches are the three types of DTI prediction techniques that can be used.

In this study, the area under the receiver operating characteristics curve (AUROC) and the area under the precision-recall curve (AUPR) were used to assess each method’s efficiency. The analysis demonstrates that NFM, a content-based recommendation system, can efficiently utilize the low-dimensional characterization from KGE and thus significantly improve the prediction Performance.

KGE\_NFM, which could be viewed as a pre-trained model based on knowledge graph and is integrated with a recommendation system tailored for a specific downstream task, captures the latent information from heterogeneous networks using KGE without any similarity matrix and then applies neural factorization machine (NFM) based on recommendation system to enforce the feature representation for a specific downstream task, which is the DTI prediction in this work

**3.METHODOLOGY**

**3.1 Precaution recommender system using supervised ML:**

****

**Fig-1: Flow diagram for module-1 & module-2**

Pre-processed data is split into training and testing sets with a ratio of 80:20. Both training and testing set pass-through feature extraction using the TF-IDF.

The extracted features from training data are used to train the machine learning models and use the 20% test data to evaluate the performance of models.

The trained model is used to predict the disease using the vice data from the user in real-time. Based on the disease prediction, precautions are recommended for the particular disease. For recommendations, a separate dataset is maintained containing the recommendations. Each row of the dataset contains the precautions for one disease.

The text data is used for pre-processing before it can be given to the trained classifier for classification. Once pre-processed, both modules have similar data. It indicates the similarity of categorical data transformed to text data and speech transformed text data and suggests the suitability of transformation for training and testing for disease prediction.

We use accuracy, precision, recall, and F1-score to evaluate and compare the performance of the proposed approach. Following are the four basic notations used in these parameters [3, 50]:

– True Positives (TP): Target class is positive, and the classifier predicts it to be positive.

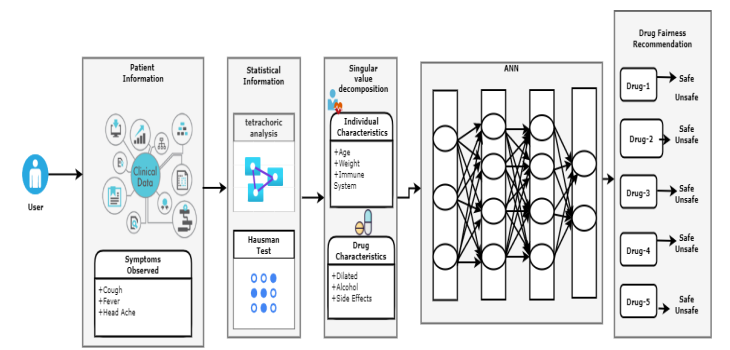
– True Negatives (TN): Target class is negative, and the classifier predicts it to be negative.

– False Positives (FP): Target class is negative but the classifier predicts it to be positive.

– False Negatives (FN): Target class is positive, but the classifier predicts it to be negative.

**3.2 Drug Recommendation System using Stacked ANN:**

A hospital recommendation system was proposed based on the treatments, consulted physicians, hospitals, and health indicators of a patient. An alternative hybrid recommender system based on available information on family doctors and available patients was suggested. The fair drug recommendation system takes into account health conditions, preferences, race, and gender. Based on the weighted binary singular value decomposition, a stacked ANN is proposed.



**Fig-2: Block diagram for the proposed recommender system.**

**ALGORITHM:**

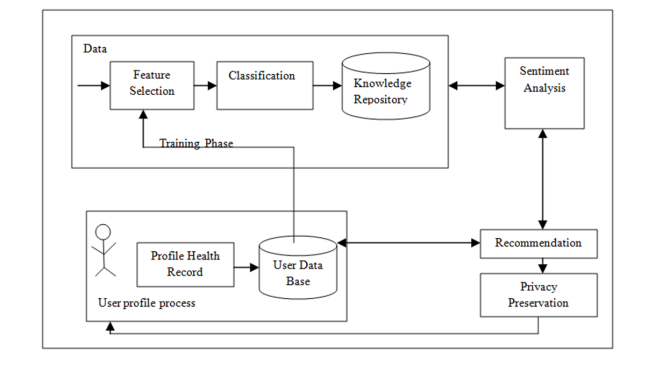
* Select the users into groups based on the filtered features, then arrange all the people who share those features together.
* Calculate the active person's medication characteristics, such as dosages, tolerances, smells, and gas production, while taking their health status and personal preferences into account.
* Generate recommendations by calculating the individual’s health status.
* To suggest drugs for the specific disease, the DRS recommends antibacterial drugs based on the individual’s past health status and present risk level.
* Generate the list of recommended drugs and dosages based on the individual’s health status, lifestyle, and individual preferences.
* If the diseased individual has allergies, high blood pressure, and poor health, adverse side effects of the drug may lead to death or morbidity. The probability of the drug side effect must be calculated.
* Based on the user’s immune system and preferences, the stacked ANN model identifies the appropriate set of medications.
  + 1. **3.3 HRS Using Collaborative Filtering:**

In recommender systems, two key actors—patients and products—play significant roles. Patients report preferences on specific items, and it is necessary to identify these preferences from the data gathered. The gathered information is displayed as a utility matrix, which gives the value of each patient-item pair, which shows the level of that patient's preferences for particular items.

The use of big data analytics has increased across a variety of fields as a result of the rapid growth of data mining and analytics. Big data analytics and its connection have established themselves in the promising field of healthcare systems,

earning their own respect and honor. Volume, velocity, and diversity are the three key defining features of big data that are present in healthcare data. Based on predictions and recommendations for the patients, the recommender system uses predictive analytics.

The proposed methodology is realistic and practical. It involves finding the objectives of the project, identifying the project`s importance, technical assessment, cost estimation, and effort estimation. We also need a framework for designing a HRS in cooperation with patients, doctors, surgeons and medical personnel.



**Fig-3: Health Recommender System (HRS).**

The following stages form up the whole framework of the health recommender system (HRS):

* Training Phase
* Patient Profile Generation
* Sentiment analysis
* Recommender
* Privacy preservation

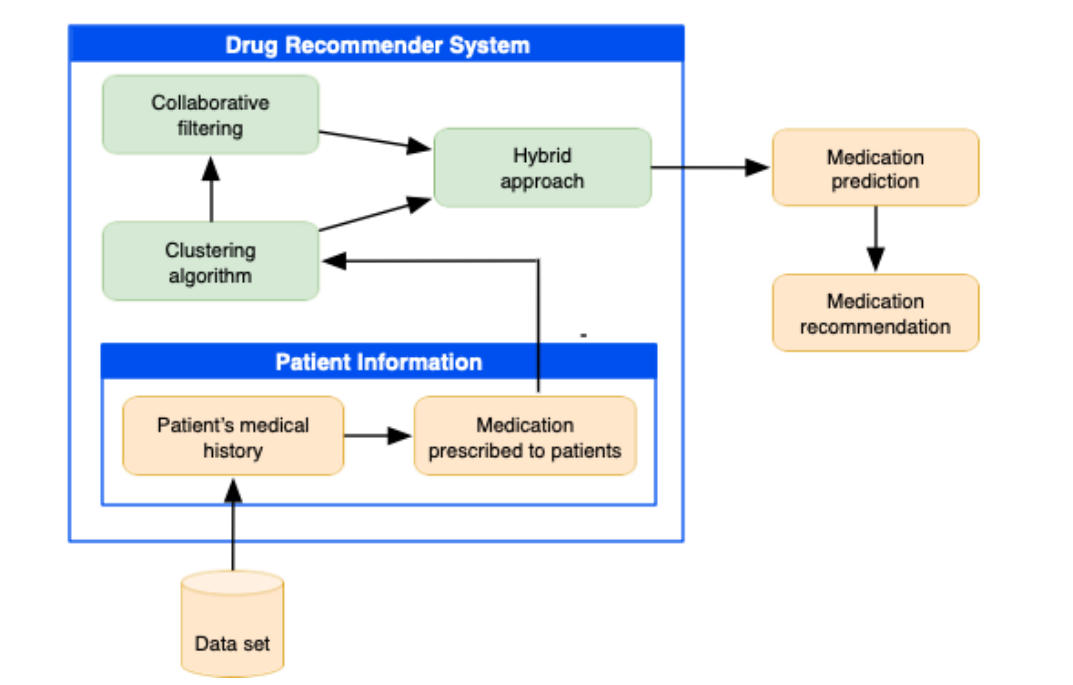
We require a framework made up of many tools that satisfies the domain requirements and the criteria of particular applications.

**3.4 DRS Using Clustering Approach:**

Exploratory analysis of the data set is a critical process in research to discover patterns, detect anomalies, test hypotheses, and test assumptions with the help of statistics and graphical representations. It is good practice to first understand the data to obtain as much information as possible. The original data set was cleansed by deleting any duplicate patient cases and records of patients who had not been prescribed any medicine before using the clustering procedures.

In addition, women had a slightly higher readmission rate than men in cases of readmission longer than 30 days. Readmission showed a similar distribution for patients with and without medication prescribed for diabetes prior to hospital treatment. In addition, we determined that the majority of the patients were of the Caucasian race and did not have a glucose or haemoglobin A1C test.

To represent the medications provided to each patient in accordance with the dose given, the proposed RS is based on the collaborative filtering approach. In order to group patients with comparable features, the clustering technique was used. The partitional K-means algorithm and the density-based spatial clustering of applications with noise (DBSCAN) technique were examined in order to identify which clustering approach produced the best patient groupings.



**Fig-4: Schematic of the proposed recommendation approach.**

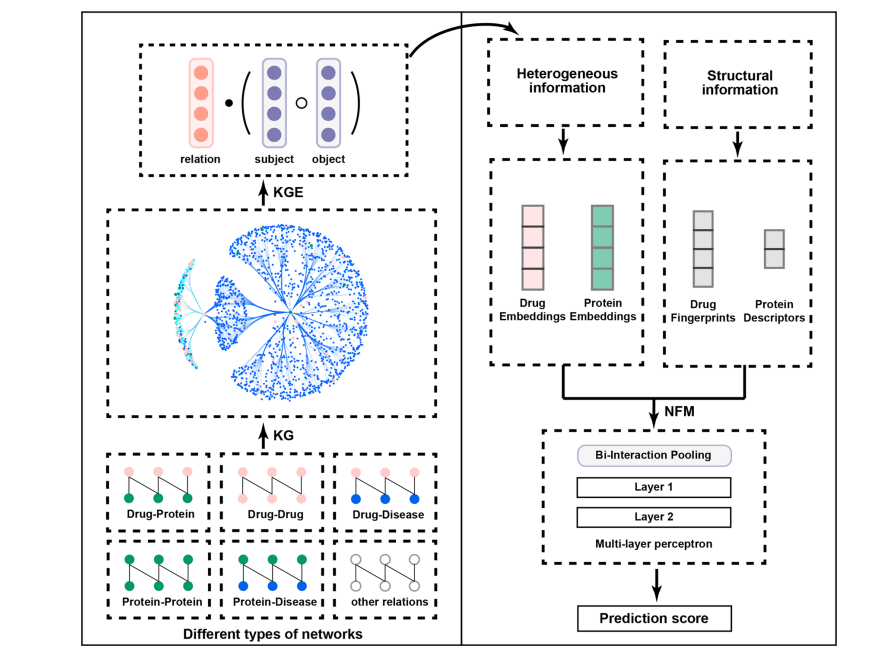
Using clustering can address several known issues in recommendation systems, including increasing the diversity, consistency, and reliability of recommendations; the data sparsity of user-preference matrices; and changes in user preferences over time. This work will be useful for both beginners in the field of recommender systems and specialists in related fields that are interested in examining the applicability of recommender systems. This review is focused on the analysis of the scientific literature on the topics of recommender systems.

**3.5 Drug-Target interaction prediction framework:**

Drug development depends heavily on the prediction of drug-target interactions (DTI) for a number of reasons, including virtual screening, drug repurposing, and the detection of potential side effects. The severe sparsity of DTI datasets and the slow response issue continue to plague existing approaches, despite significant efforts being made to improve DTI prediction. Here, we combine knowledge graph (KG) and recommendation systems to create KGE NFM, a unified framework for DTI prediction.

This type of approaches allows not only to extrapolate the prediction to discover new compounds toward known targets, but also to extrapolate the prediction to detect new targets toward known compounds.

Users can be patterned as drugs and goods can be patterned as targets for DTI predictions that make use of recommendation systems. Network-based techniques like dual regularised one-class collaborative filtering have already been merged with the widely used collaborative filtering technique for predictions. This task can be successfully completed using machine learning (ML) models, particularly network-based approaches, which have many advantages over other computational approaches.



**Fig-5: The schematic workflow of KGE\_NFM.**

In this study, we proposed a unified framework entitled KGE NFM that combines KGE and recommendation system methodologies for DTI prediction and is relevant to many drug discovery scenarios, particularly when encountering novel proteins. KGE NFM, which could be thought of as a pre-trained model based on knowledge graph and is combined with a recommendation system tailored for a particular downstream task, captures the latent information from heterogeneous networks using KGE without any similarity matrix and then applies NFM,which in this work is the DTI prediction.

**4.RESULTS AND DISCUSSION**

**4.1 Precaution recommender system using supervised ML:**

The Text to speech recognition system is used to generate test data in real-time from the spoken symptoms provided by different patients. The test data goes through the same preprocessing stages as the training data. According to the observations, textual data typically exhibits a higher level of categorization efficiency than the original categorical data. All machine learning classifiers perform extremely well enough when disease prediction is done using modified textual input.

**Table-1: Machine learning models performance on test data**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Classifier** | **Accuracy** | **Precision** | **Recall** | **F1 Score** |
| RF | 99.8% | 99.8% | 99.9% | 99.9% |
| ETC | 99.8% | 100% | 99.8% | 99.9% |
| DT | 99.8% | 99.9% | 99.9% | 99.8% |
| SVM | 99.8% | 99.9% | 99.9% | 99.8% |
| LR | 99.8% | 99.8% | 99.8% | 99.9% |
| NB | 99.8% | 99.9% | 99.9% | 99.8% |

* The results demonstrate the superiority of the pre-trained transformers over more conventional feature extraction and embedding techniques.

**4.2 Drug Recommendation System using Stacked ANN:**

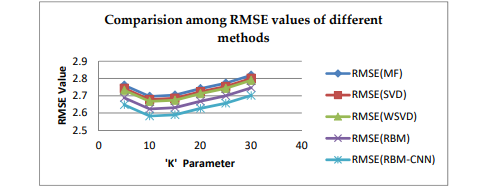
The stacked ANN was implemented for a drug recommendation system with varying hidden layers, and the error rate was compared as the number of hidden layers was increased. It was observed that the proposed model performed better than other traditional ML algorithms in terms of accuracy, precision, sensitivity, and specificity. The accuracy of the system architecture suggested in this study rapidly developed to 98.5%, showing that the ANN is the standard model for drug recommendation systems.

**Table-2: Comparison of the performance metrics of machine learning-based recommender systems.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Recommender Model** | **Accuracy** | **Precision** | **Sensitivity** | **Specificity** |
| Content-Based | 0.847 | 0.842 | 0.862 | 0.897 |
| Random forest | 0.841 | 0.840 | 0.841 | 0.920 |
| K-nearest neighbours | 0.840 | 0.823 | 0.824 | 0.915 |
| SVM | 0.719 | 0.714 | 0.719 | 0.860 |
| Logistic regression | 0.534 | 0.522 | 0.534 | 0.767 |
| Decision Tree | 0.840 | 0.840 | 0.835 | 0.920 |
| Deer Dr | 0.956 | 0.945 | 0.924 | 0.919 |
| MLP | 0.956 | 0.945 | 0.917 | 0.916 |
| Proposed | 0.985 | 0.6 | 0.939 | 0.929 |

* Based on the findings, the proposed system is effective in extracting diabetic patients’ risk factors and recommending drug therapy.
  + 1. **4.3 HRS Using Collaborative Filtering:**

Here, the experiments are conducted on data set of healthcare and it is divided into training and test data in 75:25 ratios respectively. The suggested RBM-CNN-based collaborative filtering method's RMSE changes with the variable K and is optimal at K = 10. The accuracy increases as the value decreases.



**Fig-6: Comparison among RMSE values of different methods**

* For identifying a treatment for a specific patient, the combination of an RBM with CNN in a deep learning environment increases the accuracy of the recommendations.
* This proposed method provides improved accuracy when two error metrics, such as RMSE and MAE, are taken into consideration by assessing and comparing all the approaches with the proposed RBM-CNN method.
* The combination of an RBM and CNN in a deep learning environment improves the quality of the recommendations for selecting a hospital for a particular patient

**4.4 DRS Using Clustering Approach:**

For the similar data set, the K-means algorithm had a lower number of clusters and a higher Segmentation coefficient than DBSCAN. As a result, the K-means clustering results were also taken into account when calculating predictions and suggestions.

**Table-3: Comparative analysis of the performance of the algorithms**

|  |  |  |  |
| --- | --- | --- | --- |
| **Algorithm** | **Name of clusters** | **Silhouette Coefficient** | **Execution Time** |
| K-means | 6 | 0.654 | 15min, 24sec |
| DBSCAN | 200 | 0.611 | 20 min, 31sec |

The experimental findings indicate that, when taking patient information into account, our recommendation approach performs well in terms of providing accurate forecasts and suitable recommendations. Based on these findings, an RS is required to provide support to health care professionals to facilitate the management and control of this chronic disease. Therefore, a cluster-based RS was proposed to help recommend drugs to patients with diabetes. The quality indicator is crucial for assessing the effectiveness of our RS since it tells us how many of the prescribed medications are appropriate for the user.

* Our experimental results showed acceptable performance of the proposed system.

**3.5 Drug-Target interaction prediction framework:**

In this study, we created a unified framework called KGE NFM to combine various data from many sources in order to estimate unique DTI. This enables the technique to recommend novel DTI within prior understanding of medicines and proteins, which is remarkable. KGE\_NFM was shown to be a successful pipeline for recommendation system.

Additionally, KGE NFM is a highly scalable framework that makes predictions more reliable by incorporating data from multiple modes. Additionally, we will extend our KG-based recommendation framework's field of application to include biological science.

Evaluation of the BioKG dataset's performance in three representative cases. KGE NFM was compared to six standard approaches. Using the BioKG dataset, including RF, DeepDTI, and MPNN CNN TriMode, NFM, and DistMult

KGE\_NFM proposed in this article is an efficient strategy to leverage heterogeneous data for DTI prediction. In fact, KG has tremendous potential for many downstream tasks by incorporating other algorithms in an appropriate way.

**Table-4: Summary of the baseline methods :**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Category** | **Model** | **Drug**  **featurization** | **Protein**  **Featurization** | **Heterogenous**  **information** | **Classifier** |
| End-to-end methods | MPNN\_CNN | MPNN | CNN | / | MLP |
| DeepDTI | CNN | CNN | / | MLP |
| Feature-based methods | RF | Morgan | CTD descriptors | / | RF |
| fingerprints |
| Heterogeneous data | DTINet | / | / | Network embeddings | Induce matrix completion |
| Driven methods | DTiGEMS | / | / | Graph embeddings | MLP |
| TriModel | / | / | KGE | / |
| KGE\_NFM | Morgan fingerprints | CTD desriptors | KGE | NFM |

**5.CONCLUSION**

Reviews are becoming an essential part of our daily lives; before making purchases, making an online purchase, or visiting a restaurant, we first read reviews to help us make the best choices. Results of experiments indicate that text data typically exhibit higher accuracy than category data. deep learning techniques are used to make unbiased and fair drug recommendations. By modifying the threshold value, statistical analysis is also employed to increase accuracy while balancing fairness. The accuracy of the predictions and suggestions was analyzed in order to gauge the recommender system's performance. In the future, we'll work to improve our algorithm so that it can deliver greater accuracy while maintaining a high level of privacy. Moreover, disease diagnosis varies among various medical experts for their medical experience. The suggested method employs machine learning classifiers and patient speech data to forecast disease and suggest precautions. The voice data is converted into text using the Google speech recognizer. In contrast to the traditional diagnosis procedure, the proposed approach asks the patient to describe his symptoms in order to diagnose the illness and provide any necessary preventative measures. Results of experiments indicate that text data typically exhibit higher accuracy than category data.

**6.REFERENCES**

[1].Rustam, F., Imtiaz, Z., Mehmood, A. et al*.* Automated disease diagnosis and precaution recommender system using supervised machine learning. MultimedToolsAppl (2022).

[2].Bhimavarapu, U; Chintalapudi, N.; Battineni, G. A Fair and Safe Usage Drug Recommendation System in Medical Emergencies by a Stacked ANN. Algorithms 2022,15,186.

[3].Sahoo, Abhaya & Pradhan, Chittaranjan & Barik, Dr. Rabindra & Dubey, Harishchandra. (2019). DeepReco: Deep Learning Based Health Recommender System Using Collaborative Filtering. Computation. 7. 25. 10.3390/computation7020025.

[4].Granda Morales L, Valdiviezo-Diaz P, Reátegui R, Barba-Guaman Drug Recommendation System for Diabetes Using a Collaborative Filtering and Clustering Approach: Development and Performance EvaluationJ Med Internet Res 2022;24(7): e37233

[5].Ye, Q., Hsieh, CY., Yang, Z. et al*.* A unified drug–target interaction prediction framework based on knowledge graph and recommendation system. NatCommun 12, 6775 (2021).