DISEASE PREDICTION WITH ANDROID APPLICATION

Shagun Patial [*1], Shashwat Agarwal[*2], Shruti Pathak[*3], Prabhat Verma[*4]

[*1]Krishna Engineering College AKTU, Uttar Pradesh, India.

[*2]Krishna Engineering College AKTU, Uttar Pradesh, India.

[*3]Krishna Engineering College AKTU, Uttar Pradesh, India.

[*4]Krishna Engineering College AKTU, Uttar Pradesh, India.

ABSTRACT

These days the use of Machine Learning in Medical Services is growing. The right working of all situations in our body is one of the fundamental concerns nowadays. Hence, to assemble a Disease Prediction application by considering remarkable signs and side effects and ailments through the use of exploratory data analysis and machine learning models is the objective of this exploration. This investigation is mostly centered around the wellbeing region. This examination expects to frame an indicative adaptation of the normal sicknesses dependent on the signs through the utilization of the usefulness of different information expectation strategies comprehensive of grouping in the wellbeing space. In this examination, we will do exploratory information investigation on the normal sicknesses and afterward evaluate what are most manifestations and afterward make a solid application which predicts these illnesses.

Keywords: Exploratory Data Analysis, Machine Learning, Flutter Development , Medical Services.

I.   INTRODUCTION

The medical care zone produces enormous amounts of measurements every day that can be utilized to extricate experiences for foreseeing ailment. These secret bits of knowledge in the medical services information could be utilized for powerful dynamics for an individual's wellbeing. Likewise, those districts need improvement with the guide of the utilization of enlightening records in medical services. Clinical offices should be prevalent so better choices for patient examination and treatment choices might be made. Machine Learning in medical services assists individuals with creating enormous and complex logical datasets then look at them into clinical bits of knowledge. This could likewise be employed by means of doctors in bestowing valuable clinical assets. Thus, Machine Learning when executed in medical care can cause an expansion in an individual's pleasure.

II.   LITERATURE REVIEW

Anil Prajuali made a venture "Disease Predictor [1]" which is an electronic application that predicts the illness utilizing manifestations. In his venture, he gathered informational collections from various wellbeing related locales, then, at that point he utilized the Gullible Bayes order calculation through which he had the option to get the likelihood of the illness with the given side effects.

Mr. Santhana Krishnan. J, Dr. Geetha.S had investigated "Predictions Of Heart Disease Using Machine Learning Algorithms [2]". The basic role of this exploration is to expect the potential outcomes of heart disorder in patients in expressions of rate. They utilize the heart disease dataset and to perform forecasts they did information mining order techniques that are utilized for characterizing the total dataset into two classes to be specific yes and no. Choice tree type and Innocent Bayes grouping models are utilized to characterize the illness in this examination work.

Dhiraj Dahiwade, Prof. Gajanan Patel, Prof. Ektaa Meshram have planned the "Disease Prediction Model Using Machine Learning" [3]. They amassed a patient illness dataset from the UCI AI site. For infection expectation, they analyze K-Closest Neighbor (KNN) and Convolutional neural organization (CNN) AI calculation for exact forecast of ailment and found that CNN performs better compared to the KNN Calculation in expressions of precision and time.

Vijay Shetty S, Karthik G A, M Ashwin distributed the examination paper on "Symptom Based Health Prediction using Data Mining" [4]. With the dataset of the most ordinarily displayed ailments, they built a connection to anticipating the practical ailment dependent on the entering of signs and manifestations. The proposed model

uses the ability of different AI calculations joined with normal language handling to accomplish precise forecasts. Tokenization is utilized for text preparing and is additionally mixed with uncommon calculations to test the similitudes with yields. They utilized the Choice Tree, Irregular Timberland, and Guileless Bayes Classifier calculation out of which the Innocent Bayes Calculation delivered the greatest precise outcomes.

A site called "OMGYAN" [5] contains a rundown of various illnesses and their fix by yoga. This site proposes diverse yoga asanas for various infections.

In the research paper CLASSICAL AYURVEDIC PRESCRIPTIONS FOR COMMON DISEASE [6], creators recommend ayurvedic solutions for normal infections.

III.   METHODOLOGY

Collecting the data

This is the underlying stage in a machine learning project: gathering information from a dependable source. Thus, the dataset was taken from the www.kaggle.com site in this stage. The dataset was picked depending on its attributes, for example, the way that the proportion of all the diseases is generally equivalent to one, and that it was transferred by a bona fide source, for this situation, www.kaggle.com itself. In the wake of placing the information into a data frame, you may utilize the pandas library to manage it.

Exploratory Data Analysis

This stage involves physically checking the information to confirm on the off chance that it is skilled enough for the issue proclamation. The initial step was to check whether the dataset contained any invalid qualities; since this has ostensible information, the invalid qualities couldn't be adjusted, along these lines we needed to eliminate every one of the columns that contained invalid qualities. The size of the dataset was analyzed again in the wake of dropping the invalid qualities, as though to keep away from any inclination that may have been shaped because of dropping the qualities. Following that, charts were attracted to check the main 20 side effects which are displayed in a mix of the multitude of infections.

Text Cleaning

This stage includes stripping the line of every uncommon character and digits and diminishing it to one case; it very well may be lower or capitalized. They are changed over to lowercase. The cleaned strings are saved in a list (how about we call it corpus for the present) for additional preparation.

Splitting the data

This is a basic step that should be finished following the data has been cleaned to stay away from data leakage. Data leakage implies that while the test information is being preprocessed, the preprocessor model/vectorizer ought to be totally clueless of the test information. Something else, if the preprocessor likewise gets the test information esteems, it might incline our model in true applications. Therefore, the information is parted down the middle and the 'stratify' and 'random_state' parameters are utilized. As dictated by the proportion, Stratify disseminates the interesting information esteems in an equivalent offer into the train and test. Random state guarantees that the qualities are gone into train and test in a similar arbitrary succession each time the dataset is separated, so that while hyperparameter tuning, randomized train and test may not influence the model's performance.

Text Preprocessing

The test data strings are taken care of into the Tf-Idf vectorizer to fabricate a sparse matrix in this stage. It will figure the tf-idf value of a solitary word and keep up with simply the main 5000 highlights and return the sparse matrix. After the fitting is done, the train data strings are all of a sudden changed with the assistance of this vectorizer.

Hyperparameter Tuning

This stage involves preparing the model on train data utilizing a few parameters, doing cross validation for every parameter, and choosing the ideal parameters for the to be deployed model. While tuning the parameters, the default parameters were likewise given to ensure that we didn't break it from its default state parameters while tuning the hyperparacers. A few models further developed when new parameters were added, while others performed best when their default values were utilized.

Model Training

Since we're using ensemble modelling here, the discoveries depend on three distinct models. Random Forest Classifier, Multinomial Naive Bayes Classifier, and K-Nearest Neighbor Classifier are the models utilized in this study. Thus, these models are relegated to the ideal parameters from the hyperparameter tuning stage, and the training data is hence trained on every one of them.

Model Fitting

After the training is finished, the test data is utilized to fit the model.

Model Evaluation

The test model, which currently contains the predicted values, is contrasted with the actual values and surveyed utilizing a few performance matrices. The accuracy score and confusion matrix are utilized to survey execution. On the off chance that the outcomes are agreeable, the undertaking will be finished; on the off chance that they are not, the interaction will return to the hyperparameter tuning stage, where new parameters will be given into the model, and the cycle will be rehashed until good outcomes are obtained.

Model Deployment

When the desired accuracies are met with the model then the model is passed through pickling and then it is deployed with the help of a web app and an android application.

IV.   IMPLEMENTATION

Implementing Flask API

We are using Flask to deploy our machine learning models, so that we can call our machine learning models by creating HTTP requests from anywhere.

Pickling is used to dump our trained machine learning models so that its instance can be stored in byte stream and while using these models we do not need to rerun the whole process, simply we use the predict method to make our prediction. These pickled models are going to be in flask application. We can predict disease on the basis of symptoms which can be given by the HTTP POST request. We can extract the symptoms from the request and give it to the predicted method of the trained models. Predicted Disease is returned by the models and then we send this predicted disease as a response.

This flask application is hosted in Heroku which is a cloud platform for hosting services.

"*https://aatmsahayflaskapi.herokuapp.com/*" is the http address of our hosted flask api in Heroku .

Application Development using Flutter

We are going to use Flutter for our Application Development. Flutter is a cross platform software development kit which is developed by Google. FLutter uses Dart as its primary language which has syntax like C++,JAVA. Flutter apps majorly deals with widgets, it has a wide range of pre-built widgets which can make it easier to develop applications.
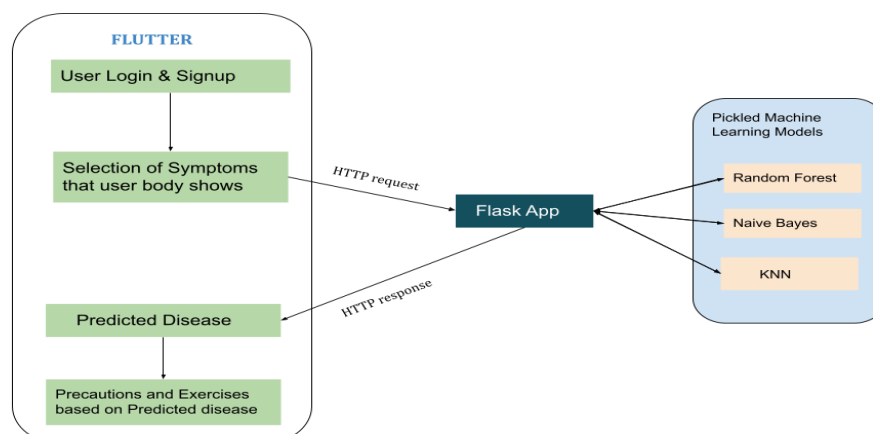
Figure 1: Workflow of Application

We are creating our app in such a way users can login and select the symptoms that his/her body shows. Then these symptoms are going to the flask web app by making HTTP post requests. Then Our flask app extracts the disease from the request and gives to the pickled machine models in the form of input to the predict method of the individual trained algorithm. Then this predict method returns the predicted disease. Flask App makes a dictionary of these predicted diseases which comes from the different machine learning models and returns an HTTP response to the Flutter application. Then this application shows the predicted disease and recommends some precautions and exercise/yoga asanas based on predicted disease.

V.   RESULTS AND DISCUSSION
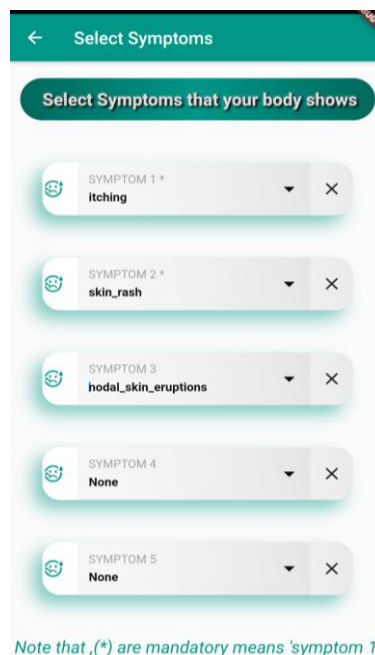
Figure 2: Selecting user Symptoms

Figure 3: Predicted disease based on the Selected Symptoms
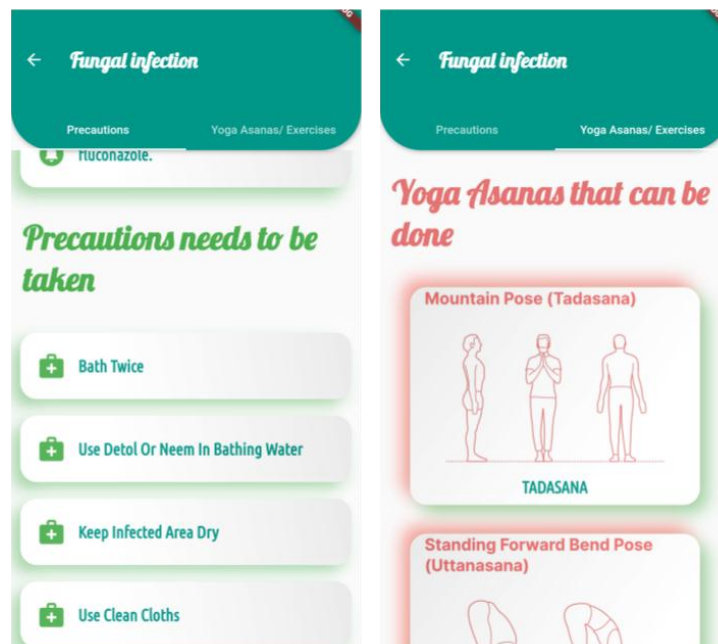
Figure 4: Precaution and Yoga asanas/ exercise based on predicted disease

The above images clearly describe how our application looks . In Fig 2, the user selects the symptoms of itching, skin rash and nodal skin eruption. Then these selected symptoms go to the flask api which will revert back the predicted disease based on that and display the disease like in Fig. 3. In Fig 4, recommendation of some precaution and Yoga asanas are shown based on the predicted disease. These results vary with the selected symptoms by users. All these results / predictions are done with the help of data we have. More the data we used to train our models gives more accurate results.

VI.   CONCLUSION AND FUTURE SCOPE

This exploration is done to achieve the higher accuracy of anticipating models, which can be used by application for expectation. This examination hopes to anticipate the sickness dependent on the indications which are given by the clients as a contribution to the application. A typical assumption exactness likelihood of 98% is gained. The Infection Anticipating calculation will send on Flask in the wake of preparation, then, at that point it is associated with the Application through the REST Programming interface. Then, at that point our application will actually want to make forecasts. The current system covers simply the most by and large happening contaminations. Future updates are expected to fuse ailments of higher loss like heart sicknesses, malignant growth, and so forth, so early estimates and meds ought to be conceivable toward the starting stage, and the setback pace of hazardous contaminants can be decreased. It can likewise be advantageous, if the application can straightforwardly show the close by accessible specific specialists for the specific anticipated illness and through which clients can straightforwardly take counteractions recommended by specialists.

VII.   REFERENCES

[1]   Parajuli, Anil. "DISEASE PREDICTOR." Ph.D. diss., Tribhuvan University, 2016.

[2]   Krishnan, Santhana, and S. Geetha. "Prediction of Heart Disease Using Machine Learning Algorithms." In 2019 1st international conference on innovations in information and communication technology (ICIICT), pp. 1-5. IEEE, 2019.

[3]   Shetty, S. Vijava, G. A. Karthik, and M. Ashwin. "Symptom Based Health Prediction using Data Mining." In 2019 International Conference on Communication and Electronics Systems (ICCES), pp. 744-749. IEEE, 2019.

[4]   Dahiwade, Dhiraj, Gajanan Patle, and Ektaa Meshram. "Designing disease prediction models using machine learning approach." In 2019 3rd International Conference on Computing Methodologies and

Communication (ICCMC), pp. 1211-1215. IEEE, 2019.

[5]   yoga.omgyan.com/cure/Disease-Yoga.html

[6]   Prof. G.S. Lavekar, Dr. M.M. Padhi, "CLASSICAL AYURVEDIC PRESCRIPTIONS FOR COMMON DISEASES"- February 2010, CENTRAL COUNCIL FOR RESEARCH IN AYURVEDA AND SIDDHA Department of AYUSH, Ministry of Health & Family Welfare.

[7]   Uddin, Shahadat, Arif Khan, Md Ekramul Hossain, and Mohammad Ali Moni. "Comparing different supervised machine learning algorithms for disease prediction." BMC medical informatics and decision making 19, no. 1 (2019): 1-16.

[8]   Chetty, Naganna, Kunwar Singh Vaisla, and Nagamma Patil. "An improved method for disease prediction employing a fuzzy approach." In 2015 Second International Conference on Advances in Computing and Communication Engineering, pp. 568-572. IEEE, 2015.

[9]   Yaganteeswarudu, Akkem. "Multi Disease Prediction Model by using Machine Learning and Flask API." In 2020 5th International Conference on Communication and Electronics Systems (ICCES), pp. 1242-1246. IEEE, 2020.

[10]   Venkatesh, R., C. Balasubramanian, and M. Kaliappan. "Development of big data predictive analytics model for disease prediction using machine learning technique." Journal of medical systems 43, no. 8 (2019): 1-8.