iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** I **Month of publication:** January 2024

DOI: https://doi.org/10.22214/ijraset.2024.58176

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue I Jan 2024- Available at www.ijraset.com

Leveraging Django and Machine Learning to Craft a Health Forecasting System, Tailoring Personalized Wellness Strategies

Proksha J Reddy[1], Dr. Sudha P[2]
[1]*PG Student,* [2]*Professor, Department of Computer science, Presidency University, Bangalore, India*
Department of Computer science, Presidency University, Bangalore, India

Abstract: ***The goal of this thesis is to develop a health forecasting system that combines Django with cutting-edge machine learning in order to customize wellness plans according to personal health information. The study addresses model selection, data preprocessing, and user interface creation as it investigates the potential synergies between machine learning and web development. Predictive health analytics potential and gaps are identified through a review of the literature. The goal of the project is to create a real-time, scalable health forecasting system while protecting user privacy. Neural networks and decision trees are two examples of machine learning methods that are utilized. The efficacy of the system in delivering precise health projections and augmenting user involvement in proactive wellness management is evaluated through empirical assessment and case studies. In order to further personalized healthcare solutions, practitioners, researchers, and developers can benefit greatly from the research's insightful contributions.***
Keywords: Health forecasting system, Django, machine learning, customize wellness plans, predictive health analytics.

I. INTRODUCTION

The amalgamation of machine learning and web development has surfaced as a revolutionary agent in the rapidly changing field of healthcare technology, bringing about unparalleled progressions in customized wellness remedies. This thesis sets out on a thorough investigation with the goal of utilizing the power of the reliable web framework Django in conjunction with state-of-the-art machine learning methods. The principal objective is to create a novel health forecasting system that utilizes personal health information to facilitate well-informed decision-making and the personalization of wellness regimens to meet the specific needs of each person.

The complex interactions between machine learning and web development are explored in this research, with a focus on important details including model selection, data preprocessing, and designing a user-friendly interface. Harmony between these components is essential for a comprehensive and successful health forecasting system to function.

Machine learning techniques are an essential component of this project. In order to extract meaningful patterns and gain understanding from the complex and ever-changing world of health data, careful selection and implementation of these algorithms are critical. Among the many methods used to examine past health data, spot trends, and forecast future health problems, decision trees, support vector machines, and neural networks stand out.

There are two steps in the machine learning portion of the procedure. Algorithms are first trained on large datasets in order to identify patterns and relationships in the data. Then, using incoming real-time data, the trained models are used to forecast and predict possible health problems. With time, the system can adjust and improve its forecast accuracy.

In addition to identifying opportunities and knowledge gaps in the area, this research attempts to provide empirical insights into the effectiveness of the integrated system by critically analyzing the body of existing literature on predictive health analytics. In order to verify the accuracy of health projections and determine the degree to which personalized wellness strategies successfully involve users in proactive health management, the study evaluates the real-time performance of the health forecasting system through empirical assessments and case studies. In the process, this research project has the potential to significantly advance the development of customized healthcare solutions and the smooth integration of online technologies and machine learning for proactive and individualized wellbeing.

II. RELATED WORK

Numerous studies have been conducted on the prediction of disease utilizing various machine learning approaches and algorithms that are applicable to healthcare facilities. This essay examines a few of the research investigations conducted using the methods in research papers and outcomes that they made use of. Below are reviews that have been submitted.

©IJRASET: All Rights are Reserved

In their study, MIN CHEN et al. [2] used machine learning techniques to create an illness prediction system. He employed methods such as the CNN-UDRP algorithm, CNN-MDRP algorithm, Naive Bayes, K-Nearest Neighbor, and Decision Tree in the prediction of disease. The accuracy of this suggested approach was 94.8%.

Sayali Ambekar et al. [3] suggested using a convolution neural network for Disease Risk Prediction. This article employs machine learning techniques such as CNNUDRP algorithm, KNN algorithm, and Naive Bayes. The system use Naïve Bayes to train it on structured data, and it achieves an accuracy of 82%.

By using a fuzzy method, Naganna Chetty et al. [4] created a system that produces better disease prediction outcomes. and employed methods such as fuzzy cmeans clustering, fuzzy KNN classifier, and KNN classifier. This study estimates the prevalence of diabetes and liver disease. The accuracy for diabetes is 97.02%, and the accuracy for liver disorder is 96.13.

Using machine learning techniques, such as CNN and KNN, Dhiraj Dahiwade et al. [5] created a model for disease prediction. In this system, the Perceptron model is utilised. Based on common symptoms such as age, sex, pulse rate, etc., this approach makes predictions about heart disease.

This proposed method has a 91% accuracy rate. A hybrid data mining and classification technique for predicting heart disease is the basis of a disease prediction system that was suggested by Ankita Dewan et al. [8]. Decision trees, neural networks, and naive bayes algorithms are some of the methods this system employs.

This system has 87% accuracy.

III. METHODOLOGY

This method primarily aims to streamline the disease prediction process, which typically requires a highly skilled physician and a significant amount of time. We have devised a technique that allows us to infer a person's illness based on his symptoms .This will be useful for people who look for assistance online, and he can consult the doctor whenever he wants. This will help you avoid hospital rush hour. When utilised independently, a variety of machine learning algorithms can provide results. But we have employed four distinct algorithms to anticipate the sickness in order to guarantee and obtain reliable results.

I have employed the following algorithms: k-Nearest Neighbours, Random Forest, Decision Tree, and Naïve Bayes. We use the classifiers stated above to forecast the disease. Several algorithms are used to provide the highest level of accuracy. The model is trained based on the comparison of the accuracies. The system makes use of this machine learning model that has been trained. Because of the increased precision, patients can consult a doctor when needed and the system becomes more efficient for them. There are roughly 5000 records in the used dataset, representing 41 diseases and 132 symptoms. After the dataset has been cleaned and reduced to prevent overfitting, algorithms are run, and the model ultimately forecasts the most likely disease.

The system's implementation details are as follows: the modules and functionalities available in the patient dashboard include: checking disease status, consulting, viewing and updating profiles, and feedback. The modules and functionalities available in the doctor dashboard include: consulting, updating profiles, and monitoring disease trends. The system is generally easy to use and comprises two users: doctors and patients. The user must first register before being able to access the offered capabilities. In the event that the user is a patient, the features that are offered include the ability to verify symptoms, access consultation records, and leave a comment. The patient has the ability to review and edit the information they submitted during registration. With the use of a machine learning model, the system makes disease predictions. Additionally, there is a way to learn more about that illness. Additionally, a doctor's list is supplied based on the anticipated disease. This will facilitate the patient's early access to the required consultation and assistance. You can accomplish this by speaking with the physician. Patients have the option to peruse the doctor's profile and make an informed decision about seeking consultations. There is now a way to view the history and specifics of consultations. It is also possible to rate things and provide comments to the system. In the event that the person registering is a physician, he can enter his information so that patients can see it. The patient's questions about the ailment are visible to the doctor, who can respond to them. The technology will serve as a bridge that connects the doctor and patient thanks to its gorgeous yet simple user interface. Utilising HTML, CSS, JavaScript, and jQuery, the UI is designed. Backend has been implemented with Django. PgMyadmin and a PostgreSQL database are also utilised.

IV. DISCUSSION

A. Models Used

1) Decision Tree Algorithm

Decision trees are thought to be a very useful and adaptable kind of classification. Pattern recognition and picture classification are two applications for it. Its great versatility makes it useful for categorization in extremely complicated problems. It can also be used to solve larger dimensions problems. Root, nodes, and leaf are its three basic components.

The attributes that have the biggest impact on the result are found in the roots, while the leaves determine the value of specific attributes and produce the tree's output. This decision tree for gain ratio is utilised. Information gain is also maximised because it employs the entropy technique. The decision tree aids in dividing the huge data set into manageable chunks. Here, input is provided in the form of symptoms (attributes) to the nodes. The output (illness) is displayed by the leaf node, while the internal nodes assist in anticipating it.
E(C)= entropy of frequency table employing one attribute is the formula used to calculate information gain: IG(C,A) = E(C)-E(C,A). With two qualities, calculate the entropy of the frequency table E(C,A). C=as-is; A=considered attribute We started our endeavour with a decision tree as our first prediction method. That provides us with about 95% accuracy.

2) Random Forest Algorithm

This machine learning algorithm is a part of the supervised learning approach. In machine learning, it addresses both classification and regression issues. This method—also known as the ensemble learning concept—combines several classifiers to address a challenging issue and enhances the model's performance. Using many decision trees on different subsets of the provided dataset, the Random Forest technique averages the results to increase the dataset's predictive accuracy. Rather than depending on a single decision tree, the random forest method uses the forecasts from each tree and predicts the outcome based on the predictions that have received the most votes. The accuracy increases with the amount of trees in the forest, avoiding the overfitting issue. The random forest algorithm's process for predicting diseases is as follows:

a) From a total of m symptoms, the algorithm randomly chooses k symptoms, and then uses these k symptoms to construct a decision tree.
b) To obtain n decision trees, step 1 is repeated several times in the next step.
c) To anticipate the sickness, pass a random variable to n decisions.
d) The most frequent projected disease is determined to be the final one after it has been calculated.

3) Naive Bayes Algorithm

The foundation of the Naive Bayes algorithm is the Bayes theorem, which presupposes the independence of the predictors. Because it doesn't require complex iterative parameter estimates, the Naive Bayes model is simple to construct and works well with enormous datasets. Notwithstanding its simplicity, the Naive Bayesian classifier is a popular algorithm because it frequently outperforms more complex classification techniques. With strong independence assumptions between the features, it applies the Bayes theorem to obtain the desired results.
The Bayes Theorem - This theorem deals with probability under certain conditions. Probability that is conditional indicates that something will occur in light of past occurrences and is dependent on past events. Given prior knowledge, the conditional probability provides the likelihood of an event. Probabilities with conditions: - P(A|B) = P(A|B).P(A) divided by P(B). In which case P(A): An alternative name for prior probability, which is the likelihood that a hypothesis is true. P(B) is the evidence's likelihood. The likelihood that the hypothesis is correct based on the evidence is expressed as P(A|B). Given that the evidence is correct, P(B|A) is the probability of the hypothesis.

4) The K-Nearest Neighbours Algorithm

This approach likewise relies on the Supervised Learning method. The K-NN method makes an educated prediction about how similar the new instance or data is to the existing cases. The new instance was then assigned to the category that most closely matched the other categories. Based on similarity, all of the available data is saved and categorised into a new data point. This indicates that the K-NN technique can be used to quickly classify newly appearing data into an appropriate category. The K-NN algorithm is mostly employed in classification-related problems. K-NN does not make any assumptions about the underlying data because it is a non-parametric algorithm. KNN is also referred to as a lazy learner algorithm since it takes time to learn from the training set.
Because the trained model of naïve bayes is the most accurate, it is utilised in the system. The ultimate outcomes together with their confidence level are shown.

V. CONCLUSION

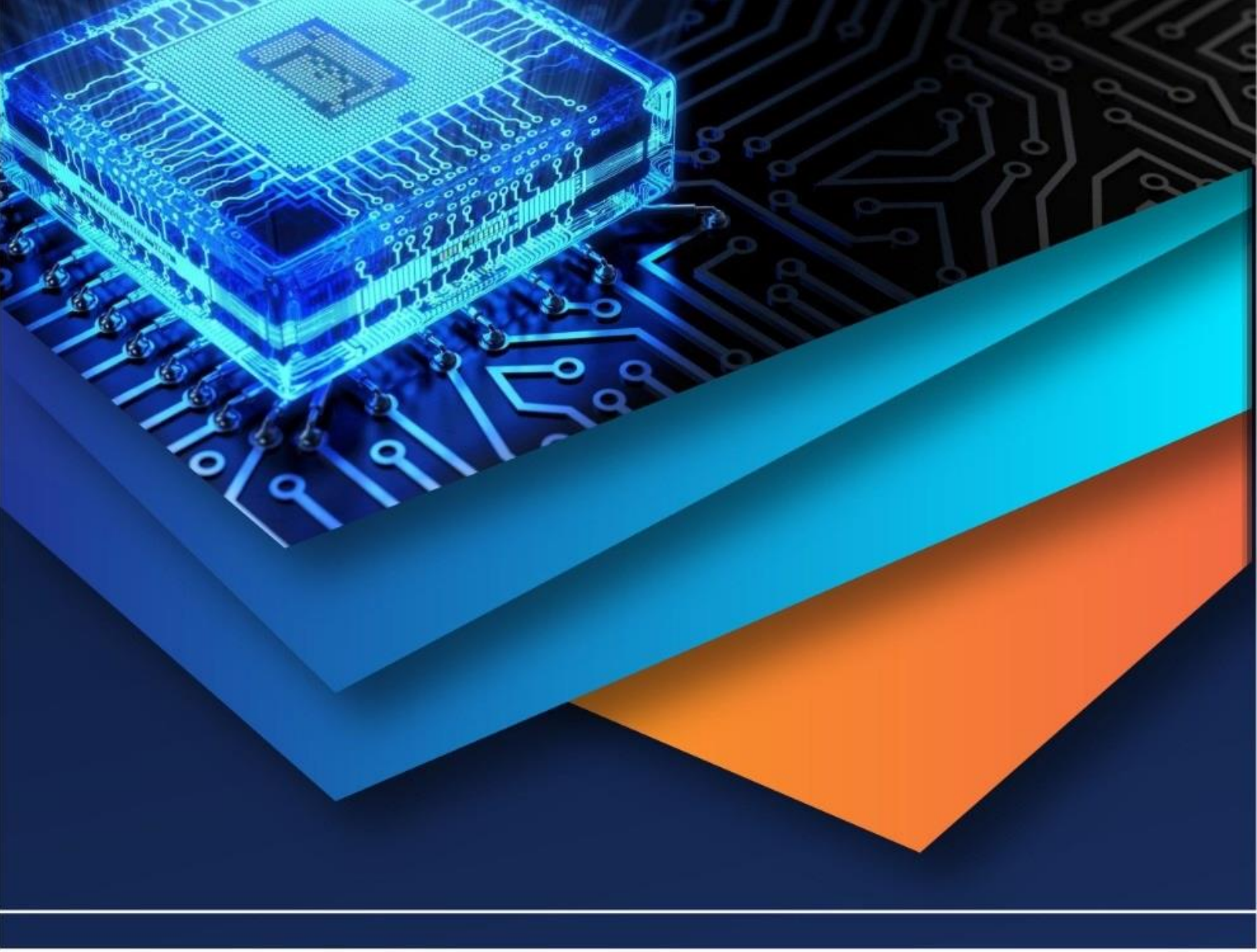
A confusion matrix, accuracy, and other metrics are used to assess the performance. The accuracy of every algorithm, including Decision Tree, Random Forest, KNN, and Naive Bayes, has been compared, and the results show that it is approximately 96%. We have thus concluded that our technique offers superior disease prediction accuracy. Comparisons indicate that Naive Bayes provides slightly higher accuracy. The model has been trained suitably, and patients who are always concerned about their health and want to know what's going on with their body will benefit from its employment in this system.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The primary goal in developing this system is to assist these individuals with their health. Additionally, small-scale physicians and clinics can use this approach to forecast disease, lessen traffic at hospital outpatient departments, and lighten the strain on medical staff. In order to benefit patients and hospitals, the makes the doctor's list of that specific predicted sickness available for immediate appointments. This guarantees patient safety, ensures that the system has no negative effects on doctors' careers, and increases the number of patients into the prediction system.

REFERENCES

[1] M. Jiang, Y. Chen, M. Liu, S. T. Rosenbloom, S. Mani, J. C. Denny, and H. Xu, "A study of machine-learning-based approaches to extract clinical entities and their assertions from discharge summaries," J. Am Med Inform Assoc, vol. 18, no. 5, pp. 601–606, 2011.
[2] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang,"Disease prediction by machine learning over big data from healthcare communities" IEEE Access, vol. 5, no.1, pp.8869–8879, 2017.
[3] Sayali Ambekar, Rashmi Phalnikar, "Disease RiskPrediction by Using Convolutional Neural Network" IEEE, 978-1-5386-5257-2/18, 2018.
[4] Naganna Chetty, Kunwar Singh Vaisla and Nagamma Patil, "An Improved Method for Disease Predictionusing Fuzzy Approach" IEEE, DOI 10.1109/ICACCE.2015.67, pp. 569-572, 2015.
[5] Dhiraj Dahiwade, Gajanan Patle and Ektaa Meshram, "Designing Disease Prediction Model Using Machine Learning Approach" IEEE Xplore Part Number: CFP19K25-ART; ISBN: 978-1-5386-7808-4, pp. 1211-1215, 2019
[6] Shahadat Uddin , Arif Khan, Md Ekramul Hossain and Mohammad Ali Moni. Comparing different supervised machine learning algorithms for disease prediction. BMC Medical Informatics and Decision Making.
[7] Ahelam Tikotikar and Mallikarjun Kodabagi. A SURVEY ON TECHNIQUE FOR PREDICTION OF DISEASE IN MEDICAL DATA. School of Computing & IT REVA UNIVERSITY
[8] M. Chen, Y. Hao, K. Hwang, L. Wang, and L.Wang,"Disease prediction by machine learning over big data from healthcare communities", ," IEEE Access, vol. 5, no. 1, pp. 8869–8879, 2017.
[9] Mr Chintan Shah,Dr. Anjali Jivani, "Comparison Of Data Mining Classification Algorithms for Breast Cancer Prediction", IEEE-31661
[10] Pingale, Kedar, et al. "Disease Prediction using Machine Learning." (2019).Mr. Chala Beyene, Prof. Pooja Kamat, "Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques", International Journal of Pure and Applied Mathematics, 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)