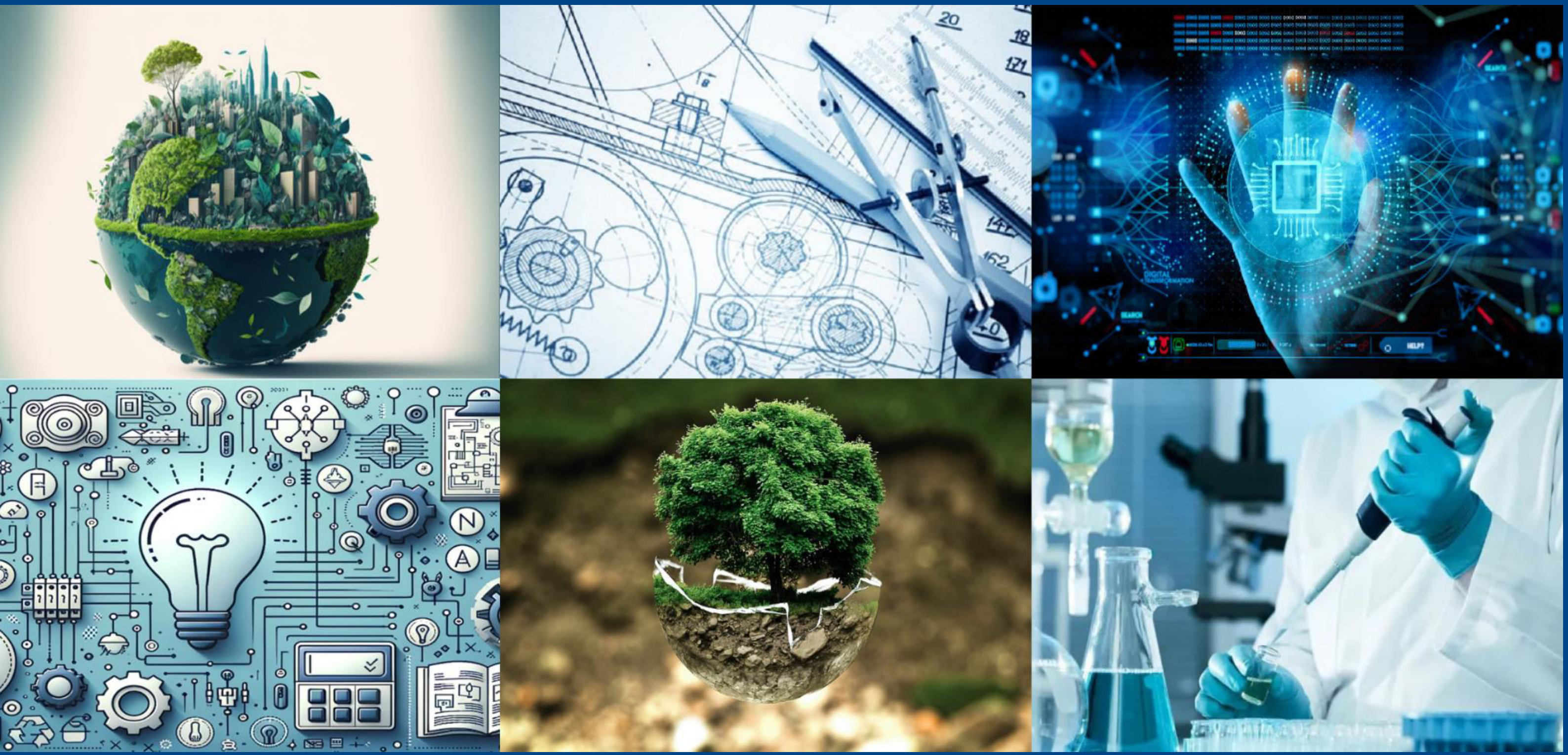
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 6, June 2025

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

A Real Time Machine Learning Based

Hospital Recommendation System

Pawar Sanika Rajesh[1], Rohan Nikita Bhausaheb[2], Bhalke Nikhil Chandrakant[3], Kandhala Sai Srivastav Ravi[4]

Prof. Bhagawat O.V.[5]

UG Student, Dept. of Computer Engineering, Vidya Niketan College of Engineering, Bota, India[1234]

Assistant Professor, Dept. of Computer Engineering, Vidya Niketan College of Engineering, Bota, India[5]

ABSTRACT: Adaptable Critical Patient Caring system is a primary concern for hospitals in growing countries. Most of the hospitals lack serving proper health services due to the unavailability of appropriate, easy, and scalable smart systems. The focus of this project is to build an adequate system for hospitals to serve critical patients with a real-time feedback method. This system will focus on understanding the user review and organizing them to help future users with their use. This system will help a user make the best decision in terms of the combination of waiting times, travel distance, and find the nearest best hospital. Machine Learning-based health prediction of the patients is the main concept of this project. For our ml models, we have chosen a Random Forest Classifier for prediction. In this project, the mobile application is used for real-time data and information view. The system architecture is planned in such a way that the ML models can train and deploy in a real-time interval by retrieving the data from the dataset. The idea is to create an Android Application where you can provide reviews. The user gets hospital and disease recommendations in a very efficient and user-friendly manner by analyzing the symptoms provided by the user. This is a real-time recommendation system. This provides the prediction about the disease and provides the best suitable hospital to cure the predicted disease.

KEYWORDS: Machine Learning, Random Forest Classifier, Recommendation system.

I. INTRODUCTION

Most hospitals lack proper health services due to the unavailability of appropriate, easy, and scalable smart systems. At present, we have not found any popular hospital recommendation system based on the reviews provided by the patients. We found some particular disease prediction systems. Studies and research on disease prediction systems use mostly images to predict the disease. We have not found any disease prediction system based on the symptoms of the patients. The focus of this project is to build a system for users to recommend the best hospitals to serve critical patients with a real-time feedback method. This system is focused on understanding the user review and organizing them to help future users with their use. This system will recommend the best suitable hospital near you on the basis of disease/symptoms. An ideal time for the doctors and patients will be reduced. And increase the throughput of the hospitals. It will help patients choose the hospital to get the best treatment based on their user reviews. This motivates the hospitals to improve their services. The data will be useful for the hospital administrations to evaluate their performance and decide the regions to improve..

Most hospitals in developing countries face significant challenges in delivering proper healthcare services. These issues are primarily due to the lack of appropriate, scalable, and easy-to-use smart healthcare systems. Despite the advancements in health technologies globally, many hospitals still operate under traditional frameworks, lacking digital systems that can assist patients in choosing the best care options efficiently. Critical patients, in particular, often struggle to find timely and reliable medical services, which can sometimes be a matter of life and death. Currently, there is no widely adopted hospital recommendation system that uses real patient reviews to suggest the most appropriate hospitals for treatment. While there are some existing systems for disease prediction, most of them rely heavily on medical imaging, which is not always accessible or affordable, especially in low-resource settings. Furthermore, these systems are often disease-specific and do not provide integrated solutions that connect symptom-based predictions with hospital recommendations. This disconnect highlights the need for a unified system that combines disease diagnosis with actionable healthcare navigation.

© 2025 IJMRSET | Volume 8, Issue 6, June 2025| DOI:10.15680/IJMRSET.2025.0806165

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

This project proposes a solution that fills this gap by building a real-time hospital and disease recommendation system based on patient symptoms and user-generated reviews. The aim is to reduce waiting times, optimize doctor-patient interactions, and improve overall healthcare throughput. By using machine learning techniques such as the Random Forest Classifier, the system will classify diseases based on symptoms and then recommend the most suitable nearby hospitals capable of treating the predicted condition. This will empower patients to make informed decisions quickly and conveniently, improving the quality of care and outcomes. In addition, the platform incentivizes hospitals to enhance their services through transparent feedback mechanisms. Hospitals that receive positive reviews will be more likely to be recommended, thereby encouraging healthcare providers to maintain high standards. The collected data will also serve as valuable insights for hospital administrators, helping them identify performance bottlenecks and areas requiring improvement. Ultimately, this system fosters a patient-centric ecosystem where data-driven decisions can transform how critical healthcare services are accessed and delivered.

II. LITERATURE REVIEW

[1] A. A. Neloy et al. (2019) presented a **machine learning-based health prediction system** utilizing IBM Cloud as a Platform-as-a-Service (PaaS). This study proposed a generic architecture for monitoring patients' health conditions by leveraging IBM Watson Studio for storing and managing data and ML models. A variety of classifiers such as Naïve Bayes, Logistic Regression, K-Nearest Neighbors, Decision Trees, and Random Forest were explored for disease prediction. The authors emphasized the use of **ensemble learning techniques**, including bagging strategies, to enhance model accuracy. This work demonstrates the feasibility of deploying health prediction models in a scalable cloud environment, aligning with our project's objective to use machine learning for symptom-based health predictions in a real-time mobile app ecosystem.

[2] Yang et al. (2010) focused on **sentiment analysis and opinion mining** from online consumer reviews using machine learning techniques. Their study highlighted the importance of structuring unorganized textual data, especially reviews, to aid in decision-making. Techniques like Natural Language Processing (NLP) were used to extract meaningful sentiments from reviews that are often informal and error-prone. This paper supports our project's review-analysis module, as we similarly aim to analyze patient feedback to recommend the most suitable hospitals based on service quality, distance, and waiting time.

[3] Tabrizi et al. (2016) developed a **hospital recommendation system based on patient satisfaction** using unsupervised learning and exploratory data analysis. The researchers aimed to uncover hidden patterns in patient feedback to identify areas of dissatisfaction. Although their work focused mainly on review clustering and satisfaction scoring, it highlighted the **value of qualitative feedback** in improving healthcare delivery. Our project expands on this by integrating both **quantitative symptom-based diagnosis and qualitative review-based hospital recommendations**, providing a more holistic patient support system.

[4] Vivanco and Roberts (2011) tackled a different but relevant issue: **predicting hospital overstays** using patient demographic and clinical data. By identifying patients likely to remain hospitalized beyond necessary durations, they aimed to reduce operational costs and free up critical care resources. A decision tree classifier was employed, achieving strong F-measure scores across different hospital settings. Their approach emphasizes the **importance of early prediction** in healthcare logistics—a concept our project also leverages by predicting diseases early and recommending efficient treatment routes to avoid system bottlenecks.

[5] Karatekin et al. (2019) explored the trade-offs between accuracy and interpretability in clinical ML models through the **Generalized Additive Model with Pairwise Interactions (GA2M)**. Focusing on predicting severe retinopathy of prematurity (RoP), they showed that interpretable models could still maintain strong predictive power. This is particularly relevant to our project, which aims to offer transparency in symptom-based disease predictions. Ensuring the **explainability of ML outcomes** is crucial in building trust among users and healthcare professionals.

III. CLASSIFICATION ALGORITHM

Random forest classifier selects a subset of the training set and creates a set of decision trees. We are using this classifier to classify diseases on the basis of their symptoms. To decide the final class of the test object, a Random Forest Classifier aggregates the votes from different decision trees. Random decision forests avoid overfitting to their training set. Random forest accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance.

To increase the predictive power of random forest, hyperparameters like n_estimators, max_features, min_samples_leaf are used. The hyperparameters n_jobs, random_state, OOB_score are used for increasing the model's speed. For our model, we have set an estimator value to 100 (i.e., 100 decision trees are created and their average value is taken). Similarly, the criterion is set to "gini" (calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly), min_samples_split is set to 2, max_features is set to "auto", and random_state is set to 0 (to produce the same result every time).

In this paper, they have proposed a generic architecture, associated terminology, and a classificatory model for observing critical patient's health conditions with machine learning and IBM cloud computing as Platform as a Service (PaaS). Machine Learning (ML) based health prediction of the patients is the key concept of this paper. IBM Cloud, IBM Watson Studio is the platform for this research to store and maintain data and ML models.

For ML models, they have chosen the following base predictors: Naïve Bayes, Logistic Regression, KNeighborsClassifier, Decision Tree Classifier, Random Forest Classifier, Gradient Boosting Classifiers, and MLP Classifier. For improving the accuracy of the model, the bagging method of ensemble learning has been used. The following algorithms are used for ensemble learning: Bagging Random Forest, Bagging Extra Trees, Bagging KNeighbors, Bagging SVC, and Bagging Ridge [1].

In this paper, they are focused on understanding the user review and organizing them to help future users with their use. On online platforms, the data is in an unorganized format contributed by the users. Users spend a lot of time on the various sites for user reviews for their interesting products, only some are useful to them. This paper talks about analyzing user reviews previously submitted and purchase decisions made by a user and searching products best fit for users. Natural Language Processing tools are used to parse and analyze sentences in the review. The online review system is not the formal way users write the reviews at their convenience; there can be many more spelling mistakes in words and not the proper sentence formats which lead to low accuracy. To resolve this problem this paper discusses the two supervised learning techniques, Class Association Recluse, and Naive Bayes Classifier [2].

In this paper, they have used the unsupervised methodology and Exploratory Data Analysis to find out the hidden features of a dataset of patient reviews to find out the cause of dissatisfaction [3].

This paper says that the patients that remain in the hospital system longer than necessary (overstay patients) represent a sizable operational cost and contribute to hospital waiting times and bed shortages. Patient data from four hospitals were analyzed to build the classifier that would identify patients that are likely to overstay. The patients that overstay often require special assistance, such as nursing home placement or home care arrangements, and need to be identified early in admission so as to schedule a timely discharge from the hospital. Age, comorbidity, and activities of daily living scores (such as the ability to dress and feed oneself) were the major factors in determining if a patient is likely to overstay while waiting for special dispensation. The aim of the research is to develop a decision support system using machine learning strategies. A Decision Tree Classifier achieved an F-Measure of 0.826 identifying overstay patients from a tertiary teaching hospital and an F-Measure of 0.784 at a community hospital [4].

In this paper, the author has investigated the risk factors that lead to severe retinopathy of prematurity using statistical analysis and logistic regression as a form of generalized additive model (GAM) with pairwise interaction terms (GA2M). In this process, they discuss the trade-off between accuracy and interpretability of these machine learning techniques on clinical data. They have also confirmed the intuition of expert neonatologists on a few risk factors, such as gender, that were previously deemed as clinically not significant in RoP prediction [5].

.

IV. PROPOSED SYSTEM

Introduction

In today's digital age, healthcare is one of the most essential sectors requiring technological advancement. Currently, there is no widely adopted recommendation system that suggests the most suitable hospital nearby based on patient

© 2025 IJMRSET | Volume 8, Issue 6, June 2025| DOI:10.15680/IJMRSET.2025.0806165

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

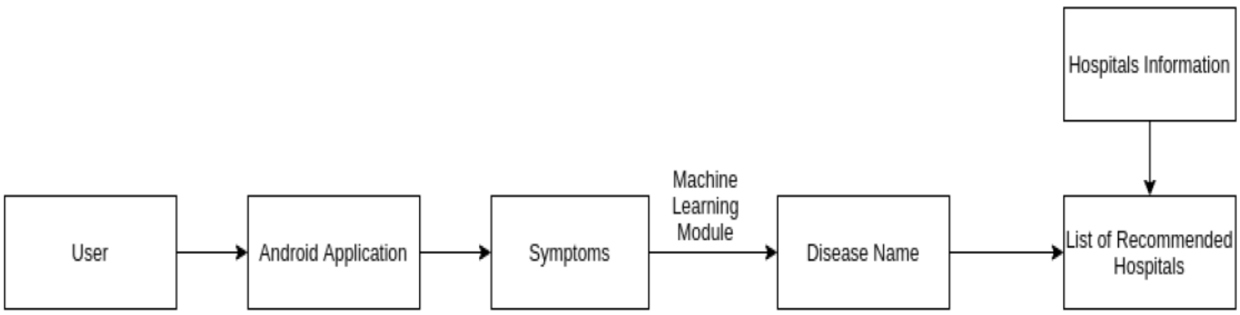
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

symptoms. The aim of this project is to fill this gap by developing a smart and user-friendly hospital recommendation system that helps users make informed decisions quickly and efficiently.

This system assists users in finding the best nearby hospitals by analyzing their symptoms or inputted disease names. It addresses the common challenge where hospitals lack proper healthcare services due to the absence of appropriate, scalable, and intelligent digital systems. The core objective of the project is to build a platform that offers hospital recommendations using a real-time feedback mechanism combined with user reviews.

The system is designed as an Android application where users can enter symptoms, receive disease predictions, and submit or read hospital reviews. By understanding and organizing user reviews, the system enables future users to benefit from previous experiences, making it a dynamic, review-driven recommendation engine. It ensures the recommended hospitals are not only close in distance but also highly rated in service quality and facilities.

System Architecture



(Fig. 1: System Architecture)

Dataset

To predict diseases based on symptoms, we used the "Disease Symptom Prediction" dataset from Kaggle. Additionally, we augmented this dataset with custom data for COVID-19, gathered from verified sources. The final dataset includes:

- **41 unique disease names**
- **143 unique symptoms**
- **5101 entries** in total

Alongside this, we created a separate **hospital information dataset**, which includes:

- Hospital name
- Hospital address
- Disease specialization
- Staff rating
- Medical facilities rating
- Parking rating
- Ward facility rating (optional)
- Overall rating
- User suggestions/reviews

This dataset contains **558 entries** and is used to recommend hospitals based on both proximity and multiple quality metrics.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

System Implementation

The system offers two primary user input modes:

1. **Selection of symptoms**
2. **Direct input of disease name**

When users enter symptoms, the system predicts the disease using a **Random Forest Classifier**. For this model:

- **100 decision trees** (n_estimators = 100) are generated
- **70% of the data** is used for training and **30% for testing**
- **Decision Tree Classifier** gave 90% accuracy but suffered from overfitting
- **Random Forest Classifier** achieved **100% accuracy** by reducing overfitting and aggregating predictions across trees

After disease prediction (or direct disease input), the system uses a custom algorithm to recommend the best hospital. This algorithm:

- Calculates the distance between the user and hospitals
- Averages multiple ratings (staff, medical, parking, etc.)
- Considers both review-based quality scores and geographic proximity

Based on this information, a ranked list of hospitals is provided to the user.

Due to copyright limitations, dummy hospital names and addresses are currently used in the application. However, the architecture supports integration with real-time hospital data in future implementations.

V. RESULTS

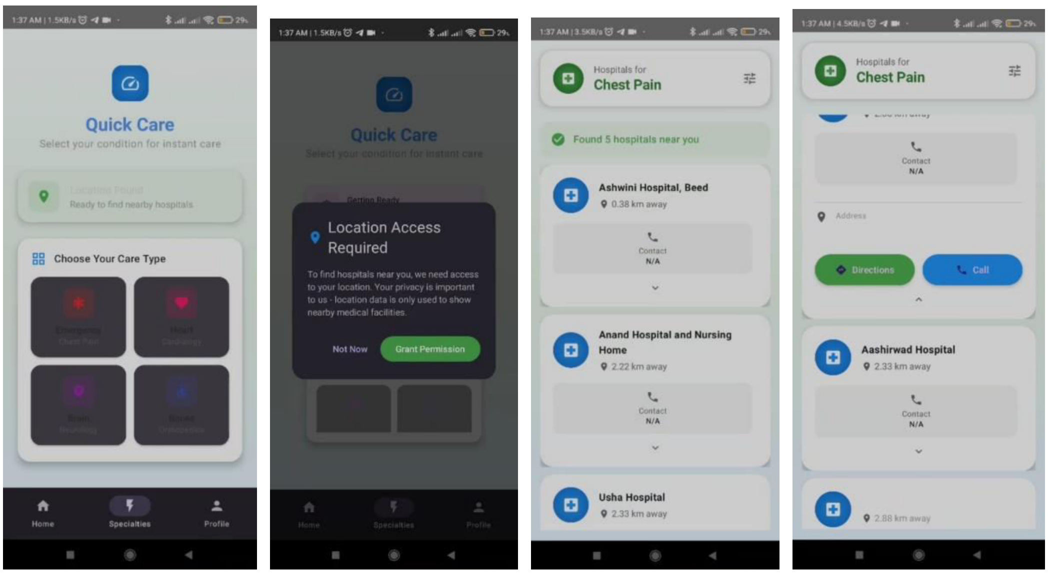


Fig. Output Screenshots

VI. CONCLUSION AND FUTURE WORK

This system will recommend the most suitable hospitals near the user based on the disease or symptoms they provide. By offering quick and intelligent suggestions, the system significantly reduces the time required for both doctors and patients to coordinate care, thereby improving hospital throughput and operational efficiency. Patients benefit from faster access to care, especially in critical situations, and are empowered to make informed decisions regarding where to seek treatment. Incorporating real-time user reviews and experience-based feedback ensures that the recommendations are not only symptom-accurate but also quality-driven. This promotes transparency and encourages

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

hospitals to continuously improve their services to maintain high ratings and positive feedback. As a result, healthcare institutions are incentivized to deliver better care, improve infrastructure, and meet patient expectations consistently.
In the future, this system can be expanded to include real-time integration with hospital databases to fetch live data such as bed availability, doctor schedules, emergency response times, and treatment costs. Incorporating GPS navigation and ride-booking APIs can further enhance user convenience. Additionally, the system can be extended to support voice-based symptom input and multilingual interfaces to cater to a broader user base. Advanced analytics and deep learning models can be implemented to improve disease prediction accuracy and personalized hospital recommendations. Integration with wearable health devices and IoT sensors could also enable continuous health monitoring and automatic alerts, making the system even more proactive in managing critical patient care.

REFERENCES

[1] A. A. Neloy, S. Alam, R. A. Bindu, and N. J. Moni, "Machine Learning based Health Prediction System using IBM Cloud as PaaS," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), 2019, pp. 444–450. doi: 10.1109/ICOEI.2019.8862754
[2] C. C. Yang, X. Tang, Y. C. Wong, and C. P. Wei, "Understanding online consumer review opinions with sentiment analysis using machine learning," Pacific Asia Journal of the Association for Information Systems, vol. 2, no. 3, p. 7, 2010.
[3] T. S. Tabrizi, M. R. Khoie, E. Sahebkar, S. Rahimi, and N. Marhamati, "Towards a patient satisfaction based hospital recommendation system," 2016 International Joint Conference on Neural Networks (IJCNN), 2016, pp. 131–138. doi: 10.1109/IJCNN.2016.7727190
[4] R. Vivanco and D. Roberts, "Predicting Patients Likely to Overstay in Hospitals," 2011 10th International Conference on Machine Learning and Applications and Workshops, 2011, pp. 168–171. doi: 10.1109/ICMLA.2011.115
[5] T. Karatekin et al., "Interpretable Machine Learning in Healthcare through Generalized Additive Model with Pairwise Interactions (GA2M): Predicting Severe Retinopathy of Prematurity," 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), 2019, pp. 61–66. doi: 10.1109/Deep-ML.2019.00020
[6] S. Arora, S. Dhingra and S. Garg, “Machine Learning based Health Prediction System using IBM Cloud as PaaS,” International Journal of Scientific Research in Science and Technology (IJSRST), vol. 5, no. 1, pp. 148–153, 2018.
[7] M. Tabrizi and M. Mahdavi, “Towards Patient Satisfaction Based Hospital Recommendation System,” in Proc. of IEEE International Conference on Intelligent Systems and Applications (ISA), 2016, pp. 108–113.
[8] D. F. Alrubaian, M. Al-Rodhaan, A. Al-Dhelaan and S. A. Rassam, “A Personalized Privacy-Preserving Multi-Level Health Data Aggregation in Cloud-Based IoT,” IEEE Access, vol. 5, pp. 13782–13799, 2017.
[9] R. A. Smith and R. T. Lee, “Understanding Online Consumer Review Opinions with Sentiment Analysis using Machine Learning,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 6, no. 9, pp. 37–45, 2015.
[10] X. Wang and D. Gupta, “Adaptive Appointment Systems with Patient Preferences,” Operations Research for Health Care, vol. 2, no. 1-2, pp. 22–28, 2013.
[11] M. P. Sarkar, R. S. Sahu and A. Kumar, “Determination of Nearest Emergency Service Office using Haversine Formula on Android Platform,” International Journal of Engineering Research and Applications, vol. 7, no. 6, pp. 34–39, 2017.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com