International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Health Management for Users from Wearable Device Data by Applying Machine Learning Algorithms

Ms. Supriya Ganesh Sapa[1], Dr. Sharvari Tamane[2]

[1]Professor, MCA, IMSCDR
[2]HOD, UDICT, MGM University

Abstract:
The growing adoption of wearable devices like fitness bands and smart watches from brands like Apple, Boat, and Noise, has enabled the continuous tracking of health parameters like physical activities – (Walking, Running) , heart rate, sleep patterns. This research focuses on how machine learning (ML) algorithms can be used to predict an individual's health status using data over 2 months from 30 individuals across various locations. The data, which varies in format depending on the device used, undergoes pre-processing to standardize and normalize the inputs, ensuring compatibility for ML analysis. Various ML algorithms, including Regression ,Decision Tree, support vector machines, and, are applied to uncover health trends, identify the health risks associated , and help in developing a personalized insights. The diverse set of participants and the extended data collection period allows for the identification of long-term health patterns and offer more accurate health predictions. The findings demonstrate the potential of leveraging wearable devices and machine learning to deliver personalized, data-driven health insights, contributing to proactive healthcare and improved fitness management.

Keywords: Wearable device, smart watch, health, track, monitor

1. INTRODUCTION

A fitness wristwatch or band can be paired with a mobile application to offer key fitness-related information and statistics to the user. Wearable devices data is used in Health Sectors to monitor the patients' health. In this research, the researcher will find how this wearable devices data can be valuable to the users of the device after analysing the data by applying Machine Learning Techniques. This will help the user in improving their health related decisions. Some ways in which wearable device data can be utilized by the user is Early Detection .This data can help to identify potential risks and encourage users to seek preventive care, reducing the likelihood of severe health issues. This encourages healthier lifestyles and promotes positive behavior change and can contribute to health improvement in the society. Real-time health monitoring is helpful in detection of health problems at an earlier stage. This enables early diagnosis, timely treatment, and better management of chronic conditions, improving health outcomes and reducing healthcare costs individuals for society.
Classification of wearable's

- Fitness and Activity Trackers- Fit bit, Garmin, Xiao Mimi Band.
- Smart watches- Apple Watch, Samsung Galaxy, Boat Watch.

- Track steps, calories burned, distance traveled, sleep patterns, heart rate.
- Health Monitoring Wearable's- Monitor heart rate, blood pressure, oxygen level, sleep patterns in elderly patients under medical supervision.
- Smart Clothing and Accessories- socks- Athletes
- Smart Glasses- Display notifications, take photos/videos.
- Hearable's
- Smart Rings- Track sleep, activity, heart rate, temperature.
- Pet Wearable's, Medical Wearable's- Patients with chronic conditions ,Kids' Wearable's ,Workplace Wearable's for Industrial workers

2. LITERATURE REVIEW

1. Using Wearables and Machine Learning to Enable Personalized Lifestyle Recommendations to Improve Blood Pressure

The paper authored by Po-Han Chiang, Melissa Wong, and Sujit Dey focuses on how wearable devices data by applying machine learning algorithms, can be helpful in providing personalized recommendations on how an individual's lifestyle should be in order to manage their blood pressure (BP).

Background:

- Common chronic diseases like High blood pressure (hypertension) is influenced by various lifestyle activities like your physical movement (Walking, Running) and your sleep duration and patterns.
- Traditional methods of managing BP through lifestyle modifications often do not account for individual differences.

Objective:

- The study aims to find how BP and lifestyle factors are related to each other and to provide personalized recommendations to improve BP using the data collected from wearable devices and applying machine learning techniques for analysing this data.

Methods:

- The proposed system collects data using home BP monitors and wearable activity trackers. It uses feature engineering to handle time-series data and improve interpretability.
- A Random Forest model with Shapley-Value-based Feature Selection (RFSV) is used to create personalized BP models and identify the most significant lifestyle factors affecting BP.
- Clinical study was done on 25 patients with elevated BP or stage I hypertension over three months to validate the system's effectiveness.

Key Components:

1. **Data Collection:**
   - BP data was collected using Omron Evolv monitors.
   - Lifestyle data, including heart rate, steps, and sleep patterns, was collected using Samsung Galaxy Watches.
2. **Feature Engineering:**
   - Data was summarized into time windows (24, 48, 72 hours) before each BP reading.
   - Features were extracted to capture temporal dependencies in BP data.
3. **Modelling and Prediction:**
   - Random Forest method was applied for BP prediction due to its strong performance with structured

data.
- Shapley values were used to interpret feature importance and provide personalized recommendations.

4. Clinical Study:
- Two groups of the participants were made, first the experimental group (received personalized recommendations) and second the control group (did not receive recommendations).
- Data related to their BP and lifestyle was collected over 90 days.

Results:
- The study validated the system's ability to provide accurate BP predictions and identify key lifestyle factors for each individual.
- Personalized recommendations based on Shapley values led to significant improvements in BP for the experimental group.
- The experimental group showed a decrease in systolic BP by 3.8 and diastolic BP by 2.3 on average, compared to 0.3 and 0.9 in the control group.

Conclusion:
- Based on data from wearable devices and machine learning Personalized recommendations on lifestyle can effectively help to manage BP.
- The study highlights the potential of using data from wearables and applying machine learning to provide precise, individualized health recommendations, improving overall health outcomes.

Practical Implications:
This study shows how personalized health interventions using wearable technology and ML algorithms can be more effective than general lifestyle advice. Health practitioners and patients can use these insights to develop tailored strategies to manage their chronic state like hypertension, leading to better health outcomes.
The paper emphasizes the importance of personalized healthcare and the role of advanced technologies in achieving it and shows a path towards more individualized and effective health management solutions.

2. Willingness to adopt wearable devices with behavioural and economic incentives by health insurance wellness programs: results of a US cross-sectional survey with multiple consumer health vignettes
The paper investigates whether the citizens of America are ready to adopt or use wearable devices for wellness programs related to health insurance and the types of incentives that might increase this willingness.

Background:
- A large number of users globally are adopting wearable devices in their daily use. These devices have potential health benefits by focusing on disease prevention, self-management, and creating long-term healthy behaviours.
- The use of wearable device data in health insurance programs could help insurers understand customer risks and needs, promote wellness, and manage claims better.

Methods:
- A survey method was applied on 997 participants from 46 states from United States of America. The participants were questioned on are they ready to use wearable devices in six different health insura-

nce scenarios, each with and without economic incentives.
- The survey included demographic questions and six hypothetical scenarios where participants had to take decision on whether to accept or reject a new health insurance service that required them to use a wearable device.

Results:
- On average, nearly 70% of participants were ready to use wearable device for their health insurance programs.
- Concerns about data security and data privacy, financial benefits and technological accuracy were common.
- Willingness to use wearables was higher for use-cases which involved more focus on health promotion and prevention of diseases.
- Economic incentives were proposed related to healthcare credits, discounts in insurance premium amount, and discounts in wellness product increased the willingness to use wearables.

Conclusions:
- Two out of every three Americans were ready to use wearable devices for their health insurance wellness programs, if the right economic, privacy, and security conditions are met.
- Financial incentives significantly improve the percentage of adoption, especially for health promotion, customised products and services based on their data, and automating the process of underwriting.

Key Points:
1. Demographics: The majority of respondents were young adults aged 25-34, predominantly female, and white.
2. Concerns: Main concerns were about the economic benefits, privacy of data, and accuracy of the data generated by the device.
3. Economic Incentives: These significantly increased the percentage of user who are ready to use wearable devices for their health insurance programs.
4. Adoption Rates: Adoption rates were high in scenarios focusing on promotion of health and prevention of diseases.
5. Implications: The study shows that if proper incentives are given then wearable-based health programs could attract more individuals from young and healthy category which will potentially stabilize the insurance premiums.

Practical Implications:

Health insurance industry could leverage wearables to promote healthier lifestyle and prevent the spread of diseases by offering financial incentives and ensuring the privacy of data and accuracy of the device. This approach could help attract and retain healthier policyholders, potentially reducing overall healthcare costs.

This review highlights that while there is a readiness to accept wearables in health insurance programs, addressing concerns related to economics, privacy, and technology is crucial for higher adoption rates.

3. Cognitive Training and Stress Detection in MCI Frail Older People Through Wearable Sensors and Machine Learning

The paper authored by Franca Delmastro, Flavio Di Martino, and Cristina Dolciotti, explores how wearable device sensors and machine learning algorithms can be used to help the patients suffering from

Mild Cognitive Impairment by providing personalized cognitive training and stress detection.

Background:

- Aging often leads to complex health conditions, including mild neurocognitive disorder, weakness and cognitive decline, which impacts the quality of life you live.
- MCI is an early stage of dementia, affecting memory and executive functions.
- Personalized training programs for cognitive and motor abilities are helpful to older people to maintain their quality of life, but these programs can increase stress levels.

Objective:

- The study aims to investigate the stress response in frail older adults during cognitive and motor rehabilitation and to develop a system for monitoring and managing stress using wearable device sensors and machine learning algorithms.

Methods:

- The study involved a pilot trial with nine frail older people with MCI. These nine people participated in cognitive and motor rehabilitation sessions.
- Wearable sensors (Zephyr BioHarness3 for heart rate and heart rate variability, and Shimmer3 GSR+ for electrodermal activity) were used to monitor physiological stress markers.
- The participants underwent cognitive training using the Stroop Color-Word Test (SCWT) and light physical exercise with a cycle-ergometer.
- The collected data was analyzed to investigate stress responses and the impact of physical exercise on cognitive performance.
- Different machine learning algorithms were used for stress detection, including Random Forest (RF) and AdaBoost (AB).

Results:

- The study showcased that physical activity improved cognitive performance in frail older adults, reducing errors and time taken to complete cognitive tasks.
- Physiological data showed an adaptive stress response, with reduced stress levels following physical exercise.
- Machine learning models, especially RF and AB, showed promising results in detecting stress with high accuracy and precision.
- Feature selection techniques helped improve model performance by identifying the most relevant physiological features.

Conclusions:

- Personalized cognitive and physical training workshops can be helpful to improve cognitive performance in frail older adults with MCI.
- Wearables and machine learning algorithms can effectively monitor and detect stress, providing valuable insights for personalizing training programs to reduce stress and improve outcomes.
- The proposed system, including a mobile app and decision support system (DSS), can support medical professionals in creating tailored interventions based on real-time stress monitoring.

Practical Implications:

This study demonstrates the potential of combining wearable technology and machine learning to provide personalized health interventions for frail older adults. By continuously monitoring physiological stress markers, the system can help optimize training programs to enhance cognitive

abilities while minimizing stress, ultimately helpe to improve the quality of life for this vulnerable population.
The paper highlights the importance of personalized healthcare and the role of advanced technologies in achieving it, offering a promising approach to managing cognitive decline and stress in older adults.

4. A Secure AI-Driven Architecture for Automated Insurance Systems: Fraud Detection and Risk Measurement

The paper authored by Najmeddine Dhieb, Hakim Ghazzai, Hichem Besbes, and Yehia Massoud proposes an advanced structure that combines artificial intelligence (AI) and blockchain technology to improve the efficiency and security of insurance operations, specifically targeting fraud detection and risk assessment.

Background:

- The insurance industry faces significant financial losses due to fraudulent claims, which lead to increased premiums and reduced competitiveness.
- Traditional methods for detecting fraud and assessing risk are time-consuming, inaccurate, and heavily reliant on human intervention.

Objective:

- The study focuses on the development of a secure, automated insurance system that reduces human interaction, improves fraud detection, and measures risk more accurately using AI and blockchain technologies.

Methods:

1. **Blockchain Integration:**
   - The framework employs a permissioned blockchain to ensure secure transactions and data sharing among different entities within the insurance network.
   - Smart contracts are used to codify business rules, automate claims processing, and validate transactions.
2. **AI and Machine Learning:**
   - The extreme gradient boosting (XGBoost) algorithm is used for batch learning to detect and classify fraudulent claims and predict customer risk.
   - An online learning solution using the Very Fast Decision Tree (VFDT) algorithm dynamically updates the model with real-time data without retraining from scratch.
3. **Data Collection and Processing:**
   - Data is collected from various sources, including customer information and claims records, and stored on the blockchain.
   - The data is cleansed, explored, and anonymized to preserve privacy before being used to train the machine learning models.

Results:

- The XGBoost algorithm achieved higher accuracy in detecting fraudulent claims (7% higher than decision tree models) and predicting future claim amounts.
- The VFDT algorithm demonstrated superior performance in online learning scenarios, effectively updating the model with new data and maintaining high accuracy.
- The proposed system successfully integrated AI and blockchain technologies to create a secure, efficient, and automated insurance framework.

Conclusion:

- The developed architecture, SISBAR, significantly improves the detection of fraudulent claims and the measurement of risk in the insurance industry.
- By reducing human intervention and leveraging advanced technologies, the system enhances the overall efficiency and security of insurance operations.

Practical Implications:

This study highlights the potential of combining AI and blockchain technologies to revolutionize the insurance industry. The proposed system can help insurance companies reduce fraudulent claims, improve risk assessment, and lower operational costs, ultimately benefiting both insurers and policyholders.

The paper emphasizes the importance of adopting cutting-edge technologies to address longstanding challenges in the insurance sector, offering a promising solution for enhancing fraud detection and risk management.

5. Health Risk Prediction By Machine Learning Over Data Analytics

The authors Prabhu T., Darshana J., Dharani Kumar M., and Hansaa Nazreen M explore in "Health Risk Prediction by Machine Learning over Data Analytics" how machine learning and data analytics can forecast health risks targeting chronic diseases. The analysis combines structured along with unstructured data systems to improve disease risk prediction models.

Background:

- Chronic diseases are a significant health and economic burden, with a large portion of healthcare costs dedicated to their treatment.
- Early detection and preventative measures are crucial to manage and reduce the impact of chronic diseases.

Objective:

- The paper proposes a Collaborative Assessment and Recommendation system that uses a combination of a patient's medical history, lifestyle habits, and big data from similar records to predict disease risks.
- The study introduces a Convolutional Neural Network-Based Multimodal Disease Risk Prediction (CNN-MDRP) algorithm that utilizes both structured and unstructured data from hospitals to improve prediction accuracy.

Methods:

1. **Data Collection:**
   - The dataset consists of electronic health records combined with medical images as well as genetic data collected from a central Chinese hospital that monitored a patient population of 31,919 with 20 million records spanning 2015 to 2018.
   - Patient information categorizes into three main groups where laboratory results and patient demographics together with lifestyle habits represent structured data.
   - Unstructured data consists of textual information such as patient narratives, doctor notes, and diagnostic records.
2. **Data Processing:**
   - Latin factor models within this study serve to estimate medical record data points that are absent from the database.

- Hospital specialists use structured data features extraction while the CNN algorithm selects features from unstructured data.

3. CNN-MDRP Algorithm:

- Structured Data Processing: Traditional machine learning methods like Naive Bayes (NB), K-Nearest Neighbors (KNN), and Decision Tree (DT) are used to predict disease risks based on structured data.
- Text Data Processing: A CNN-based unimodal disease risk prediction (CNN-UDRP) algorithm processes medical text data.
- Multimodal Prediction: The CNN-MDRP algorithm integrates structured and unstructured data to provide a comprehensive disease risk prediction model.

Results:

- The CNN-MDRP algorithm achieved higher prediction accuracy (94.8%) and faster convergence compared to the CNN-UDRP algorithm.
- The performance of CNN-MDRP in predicting disease risks is superior to traditional methods, particularly for complex diseases like cerebral infarction.

Evaluation:

- The study uses metrics such as true positive rate (TPR), false positive rate (FPR), receiver operating characteristic (ROC) curve, and area under the curve (AUC) to evaluate the model's performance.
- The accuracy of risk predictions depends on the diverse hospital data features because structured and unstructured data sources need to be integrated.

Practical Implications:

The study proves the successful implementation of advanced machine learning and big data analytical methods for risk prediction in health. The proposed system improves its risk assessment accuracy by processing structured alongside unstructured data information. The method provides early warning of medical conditions together with preventive healthcare thus it lowers healthcare costs while creating better patient results.

The paper highlights the importance of utilizing diverse data sources and advanced algorithms to enhance predictive modeling in healthcare, offering a promising solution for managing chronic diseases through early intervention and personalized treatment strategies.

6. Machine Learning for Healthcare Wearable Devices: The Big Picture

The research paper "Machine Learning for Healthcare Wearable Devices: The Big Picture" by Farida Sabry et al gives an extensive examination of current healthcare wearable device practices and their implementation challenges. This research examines how wearable devices perform both clinical disease diagnosis and elderly care support and patient monitoring activities while tracking vital signs and human movements.

Background:

- Research about the integration of artificial intelligence (AI) and machine learning (ML) in healthcare applications has gained extensive focus throughout recent years.
- Healthcare applications receive important medical information from continuous signal monitoring through Internet of Medical Things (IoMT) devices called wearables.

Objective:

- The primary objective of this paper includes reviewing modern ML research used for healthcare wearable devices while analyzing related challenges followed by solutions derived from published literature.

Methods:

- The paper examines recent ML research on wearable devices which was published between 2017 and December 2021 through categorization.
- The paper examines different challenges ML applications face when deployed on wearable devices by discussing service deployments along with power usage and user acceptance rates along with reliability and security measures.

Key Findings:

1. **Wearable Devices and Human Signals:**
   - The combination of wearable sensors functions to track human physiological together with psychological states.
   - The devices utilize multiple sensors such as skin temperature alongside EDA as well as ECG and EEG and EMG and PPG together with accelerometers gyroscopes and magnetometers.
2. **Applications of ML in IoMT:**
   - Machine learning plays a crucial role in the Internet of Medical Things (IoMT) by enabling applications such as activity recognition, fall detection, stress monitoring, fitness tracking, health parameter monitoring, and disease diagnosis.
   - Researchers have conducted considerable work on ML applications but few such innovations have become commercially available for wearables.
3. **Challenges:**
   - Data Availability and Reliability: The success of medical ML models depends on obtaining consistent high-quality data resources. The task of maintaining accurate sensor data consistency proves difficult to accomplish.
   - Model Selection and Reliability: The selection process for ML models requires the selection of ML models depends on achievement accuracy while ensuring interpretability with appropriate computational demands and memory needs.
   - Deployment Alternatives: The deployment of ML models exists for wearable devices as well as edge devices and cloud systems which present trade-offs between performance speed and energy usage and security requirements.
   - Power Consumption: Wearable devices have limited battery life, and ML computations can be power-intensive.
   - Storage and Memory: Limited memory on wearable devices requires efficient data handling and model optimization techniques.
   - Utility and User Acceptance: User acceptance of wearable devices depends on their perceived benefits, comfort, and privacy concerns.
   - Security and Privacy: Protecting personal data and ensuring secure communication is critical for user trust.
   - Potential Solutions:
   - Data Compression and Dimensionality Reduction: Lowering the processing data volume helps organizations handle their power usage and storage capacity limitations.
   - Model Compression and Optimization: Techniques such as pruning, quantization, and using efficient ML models can make them suitable for deployment on wearable devices.

- Edge Computing: Offloading computations to edge devices can balance the load and enhance real-time processing capabilities.
- Privacy-Preserving Techniques: Implementing methods like federated learning can help keep user data private and secure while still benefiting from ML insights.

Conclusion:

The research presents both the major healthcare advantages of ML and wearable devices and details the technical as well as practical implementation barriers. The future success of incorporating ML with wearable devices depends on solving technical and practical challenges that will enable revolutionary healthcare monitoring and diagnostics for better health outcomes.

This paper delivers important findings about modern ML technology applied to wearable devices along with future research guidelines which stress continuous development to maximize healthcare benefits from these technologies.

7. Survey on: Applications of Smart Wearable Technology in Health Insurance

The paper titled "Survey on Applications of Smart Wearable Technology in Health Insurance" by Apeksha Shah, Dr. Swati Ahirrao, Dr. Shraddha Phansalkar, and Dr. Ketan Kotecha provides an in-depth review of how smart wearable technology is being integrated into the health insurance sector. It discusses the benefits, applications, challenges, and potential solutions for leveraging wearable technology to improve health outcomes and optimize insurance operations.

Background:

- Present-day wearable devices have become popular among users because they enable the tracking of health metrics including heart rate and blood pressure together with step count and BMI along with calorie measurement.
- Health monitoring devices provide individuals with health assessment capabilities which deliver crucial information for insurance companies to control risks and enhance service delivery.

Objective:

- The study aims to examine how wearable technology, combined with AI, can be used to predict health status and its outcomes.
- It explores the benefits of integrating wearables in health insurance, such as incentivizing healthier lifestyles and improving claim cost predictions.

Methods:

- The paper reviews various studies and use cases of wearable technology in healthcare and health insurance.
- It analyzes the applications of AI and machine learning in processing data from wearables to predict health risks and improve insurance operations.

Key Findings:

1. Applications of Wearables in Health Monitoring:

- Wearable devices can track a wide range of health metrics and provide real-time alerts about an individual's health status.
- The gathering of real-time data using these devices enables better predictions of claim costs while offering health-related rewards to customers through wellness programs.

2. AI in Wearable Technology:

- Artificial Intelligence (AI) and machine learning techniques, including Convolutional Neural Netwo-

rks (CNN) and Support Vector Machines (SVM), are employed to process and interpret data collected from wearable devices.
- These technologies can predict health risks, customer retention rates, and detect fraudulent claims.

3. Use Cases in Health Insurance:
- Health Data Acquisition: Wearables collect vital health data, improving the accuracy of claim cost predictions.
- Incentivization: Insurance companies can offer rewards and discounts to encourage healthier behaviors.
- Health Monitoring: Wearables provide real-time health status updates and alerts.
- Health Predictions: AI models can predict potential health issues, improving customer retention and fraud detection.
- Competitive Position: Using wearables strengthens the competitive position of insurance companies in the market.

4. Challenges and Limitations:
- Data Privacy and Security: Concerns about data privacy and security can hinder the adoption of wearable technology.
- Accuracy and Reliability: The accuracy of wearables in measuring health metrics is crucial for reliable predictions.
- User Acceptance: Surprisingly many people avoid disclosing their health information to insurance institutions.

Conclusion:
- Smart wearable devices connected with AI systems allow health insurance providers to optimize both patient health tracking and forecasting risks and maintain better client relationships.
- The data privacy concerns together with accuracy and user acceptance challenges can be resolved through incentivization policies backed by sophisticated data security systems.
- The paper emphasizes the potential for wearable technology to transform the health insurance industry by providing personalized, real-time health insights and improving overall efficiency and customer satisfaction.

Practical Implications:
Wearable technology adoption in health insurance demonstrates its crucial role to advance predictions and stimulate healthy conduct and shrink operational expenses. Insurance companies will provide better services through AI implementation to gain market leadership when they solve their existing challenges. The research indicates positive potential for wearable technology implementations in health insurance that will deliver mutual advantages to insurers and their policyholders.

3. OBJECTIVE
- To study the impact of health tracking using wearable devices to users for their health management..

4. RESEARCH METHODOLOGY
The analysis takes place with different ML methods including Linear Regression, Multiple Linear Regression, k-means clustering, Random Forest classifier for regression, decision tree for regression. The data analysis utilizes information available on Kaggle storage platform. Wearable devices provided the data included in the dataset.

Dataset has 2 files

- heartrate_seconds_merged.csv
- dailyActivity_merged.csv

A. Research Design: The research design is Exploratory research design.

B. Data Collection

The consumed Fitbit dataset running from 03.12.2016 to 05.12.2016 constituted the research material employed by Brinton J, Keating M, Ortiz A, Evenson K, Furberg R in their study titled "Establishing Linkages Between Distributed Survey Responses and Consumer Wearable Device Datasets: A Pilot Protocol." Their main objective focused on developing remote data collection methods alongside procedures to link survey self-report information to wearable device biometric data to produce an anonymized dataset. Analysis of the data involves secondary information from 30* individuals who used Fitbit devices for two months.

C. Statistical tools for analysis

The researcher employed the chi-square test for statistical analysis. Descriptive statistics, including frequency, relative frequency, mean, mode, proportions, percentages, bar graphs, and pie charts, were used as needed

Machine Learning Techniques like - Regression, Decision Tree, Random Forests, Clustering Techniques etc . are used . Principal Component Analysis -dimensionality reduction technique is used to simplify the data.

Programming Language used - R programming.

The statistical tools such as Line Chart, Bar Diagram and Box Plot will be used for data presentations.

D. The analytical data relies on secondary information because achieving primary data would be challenging due to the diverse range of brand watches used by users. Hooking data depends on various methods between different brands. There exist certain brands which prohibit their users from exporting their data. The data format for different brands exists in separate structures. Each brand employs different characteristics in their products.

E. Data Processing

heartrate_seconds_merged

Fields- Id Time Value

d %>% group_by(Id) %>% summarise(Value= median(Value))

df=dailyActivity_merged

merged_inner <- inner_join(df, d, by = c("Id", "ActivityDate"))

- Id ActivityDate TotalDistance SedentaryMinutes Calories HeartRate
- # Normalize the data
- normalize <- function(x, min, max) { return ((x - min) / (max - min))}
- # Normalized values
- normalized_distance <- normalize(TotalDistance, 0, 10)
- normalized_heart_rate <- normalize(HeartRate, 60, 180)
- normalized_calories <- normalize(Calories, 0, 5000)
- normalized_sedentary_minutes <- normalize(1440 - SedentaryMinutes, 0, 1440)
- # Invert sedentary minutes
- # Weights weight_distance <- 0.25
- weight_heart_rate <- 0.25

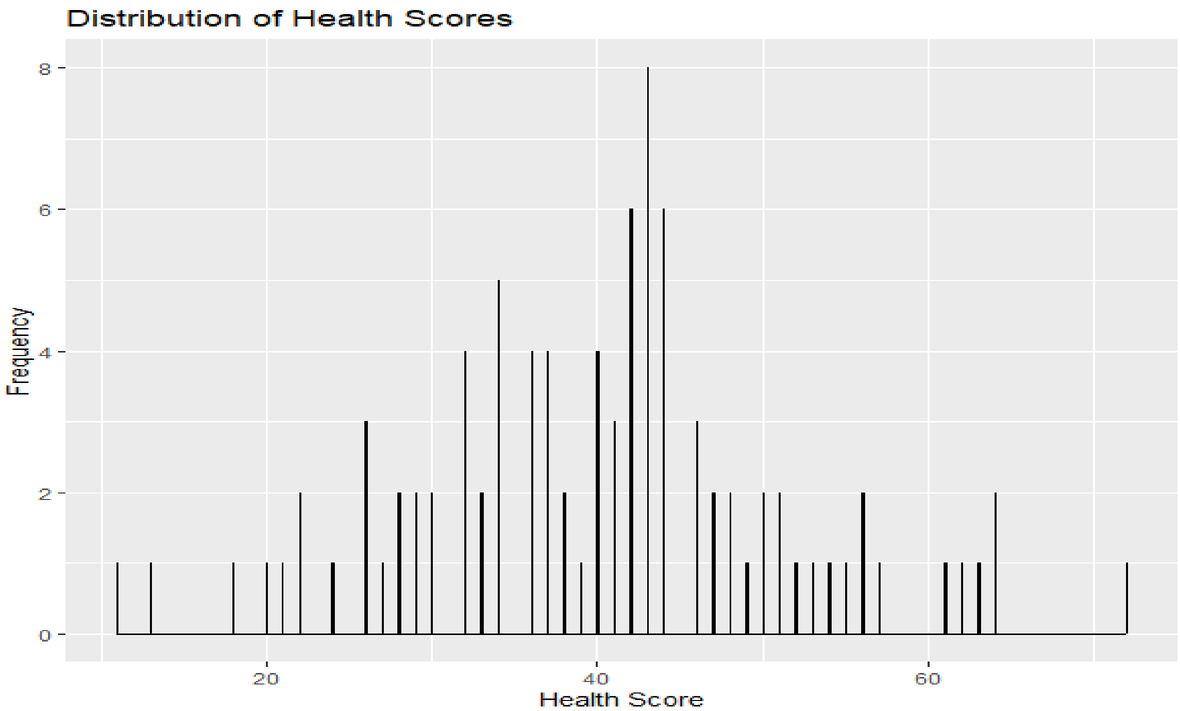
- weight_calories <- 0.25
- weight_sedentary_minutes <- 0.25
- # Calculate health score
- health_score <- (normalized_distance * weight_distance) + (normalized_heart_rate * weight_heart_rate) + (normalized_calories * weight_calories) + (normalized_sedentary_minutes * weight_sedentary_minutes)
- final = final %>%mutate(health_score)

Suggestion by doctors
- Calories Burnt depends on body weight.
- More weight- more calories burn.
- Less Sleep – Anxiety, link between insufficient sleep and increased levels of anxiety
- Less sleep – low heart rate
- During deep sleep stages, the heart rate naturally slows down, allowing the heart to rest and replenish its energy reserves.
- During Exercise – high heart rate

After consulting the doctors ,health score was categorized into 3 groups
- Health score – below 40 - poor health
- 41 to 60 - fair health
- 61 and above - Good health

F. Exploratory Data Analysis (EDA)

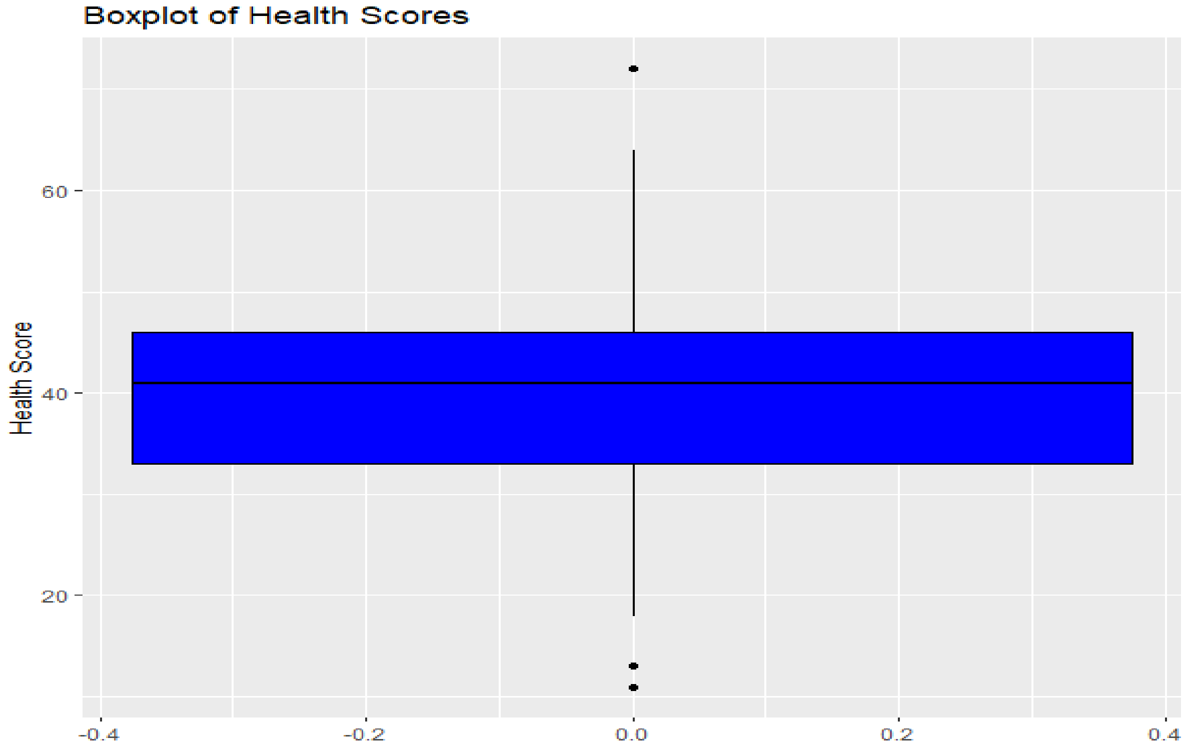
Histogram of health scores



Interpretation:

The data distribution shows its peak occurrence at score 40 which indicates most of the data points cluster in that range. Some health scores are much lower (around 10–20) or higher (around 60–65), which might indicate outliers or less common cases. The spread of the data is fairly wide, indicating variability in health scores.

Boxplot of health scores



Interpretation

The dots below and above the whiskers indicate outliers in the data. Extremely deviant values exist outside a distance of 1.5 times the IQR both beneath Q1 and above Q3. There are a few low outliers (below 20) and possibly one high outlier (above 60).

Correlation

correlation_matrix <- cor(final[, c("TotalDistance", "HeartRate", "Calories", "SedentaryMinutes", "health_score")])

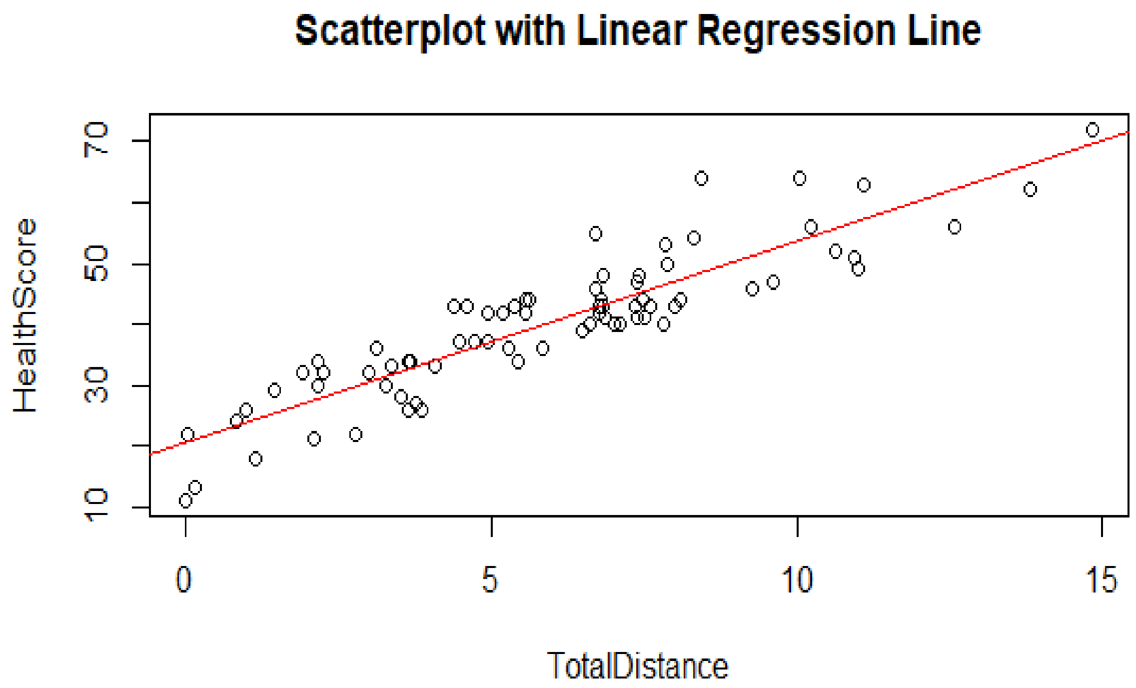
print(correlation_matrix)

	TD	HR	Cal	SedMin	health_score
TD	1.00000000	0.1672768	0.5192513	-0.05568553	0.8963220
HR	0.16727679	1.0000000	0.2714185	0.47432101	0.2051977
Cal	0.51925130	0.2714185	1.0000000	0.14880447	0.6695221
SMin	-0.05568553	0.4743210	0.1488045	1.00000000	-0.3204227
h_score	0.89632202	0.2051977	0.6695221	-0.32042267	1.0000000

Interpretation:

- Increased physical activity (TotalDistance) and higher calorie expenditure (Calories) are strongly associated with a better health score.
- More sedentary time (SedentaryMinutes) is negatively associated with health score.
- The correlation between heart rate and other factors is moderate yet the direct effect of heart rate on health score remains lower than that of distance and calorie metrics.

G. Data Analysis and Findings -
Linear Regression HealthScore ~ TotalDistance,



Output Interpretation:

- Y=mx+c
- HealthScore =slope(TotalDistance)+ Intercept
- HealthScore = 20.582 (TotalDistance)+ 3.309
- All p-values measure below 2e-16 thus proving both coefficients to be extremely important factors.
- Information from the statistical model demonstrates that TotalDistance relates significantly with HealthScore in an effective manner.
- # predict HealthScore
- correlation_accuracy <- cor(actuals_preds)
- > correlation_accuracy
- actuals predicteds
- actuals 1.0000000 0.9355821
- predicteds 0.9355821 1.0000000

DATA EXPLORING

- Built LR model similarly for all other attributes - TotalDistance , SedentaryMinutes , Calories, HeartRate
- A correlation matrix serves as a beneficial tool which enables model developers to identify insignificant features that should be removed from their model.
- TotalDistance , SedentaryMinutes , Calories, HeartRate - these variables have strong relationship with healthscore

Multiple Linear Regression HealthScore ~ TotalDistance + Calories + HeartRate + SedentaryMinutes

Interpretation:

- HealthScore=12.76+2.497×TotalDistance+0.005018×Calories+0.2019×HeartRate−0.01719×SedentaryMinutes
- The very high F-statistic and the associated p-value (< 2.2e-16) suggest that the model is highly significant.
- This model seems to fit the data extremely well, with a very strong relationship between HealthScore and the predictors (TotalDistance, Calories, HeartRate, SedentaryMinutes).
- Each individual's 2 months data is given to train the model.
- Objective is fulfilled - The model provides personalized lifestyle recommendations to help individuals to manage their health.

Decision Tree -Regression problem

health_score ~ TotalDistance+HeartRate+Calories+SedentaryMinutes

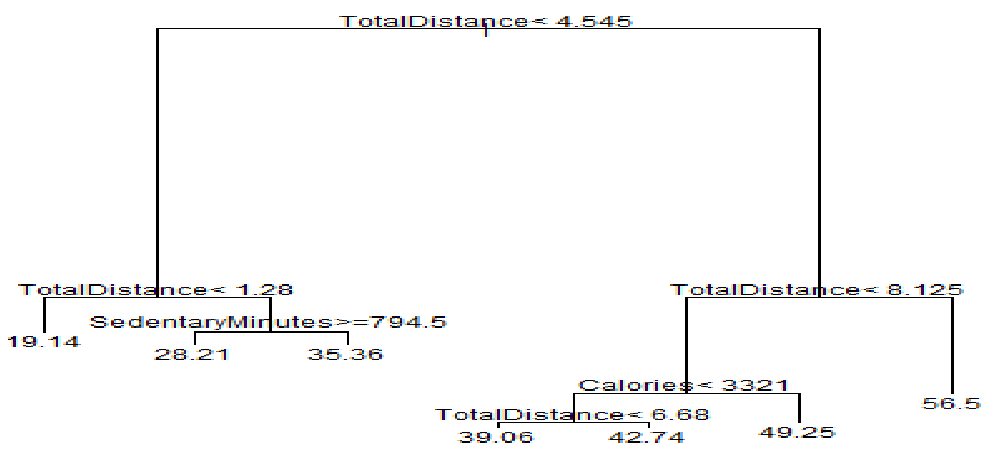
The regression model performance metrics include the following in regression applications:

- Mean Absolute Error (MAE): 3.227326
- Mean Squared Error (MSE): 17.84658
- R-squared (R^2): 0.863907

Interpretation of the Results

- MAE (3.227326): On average, the predicted values are off by about 3.23 units from the actual values.
- MSE (17.84658): The MSE ratio evaluates prediction accuracy through the medium of predicted value differences squared. Larger errors receive greater penalties from MSE than they do from MAE.
- R^2 (0.863907): The predicted values successfully account for 86.39% of the actual value variance.
- The strong model-data match is indicated by these results.

Decision Tree



Interpretation:

- If a person's TotalDistance is less than 1.28, their predicted outcome is:

- 19.14, if SedentaryMinutes < 794.5.
- 28.21, if SedentaryMinutes >= 794.5.
  - If a person's TotalDistance is between 4.545 and 8.125, and their Calories < 3321, the predicted value depends on whether their TotalDistance is also less than 6.68.
  - If a person's TotalDistance is greater than 8.125, the predicted outcome is 56.5.

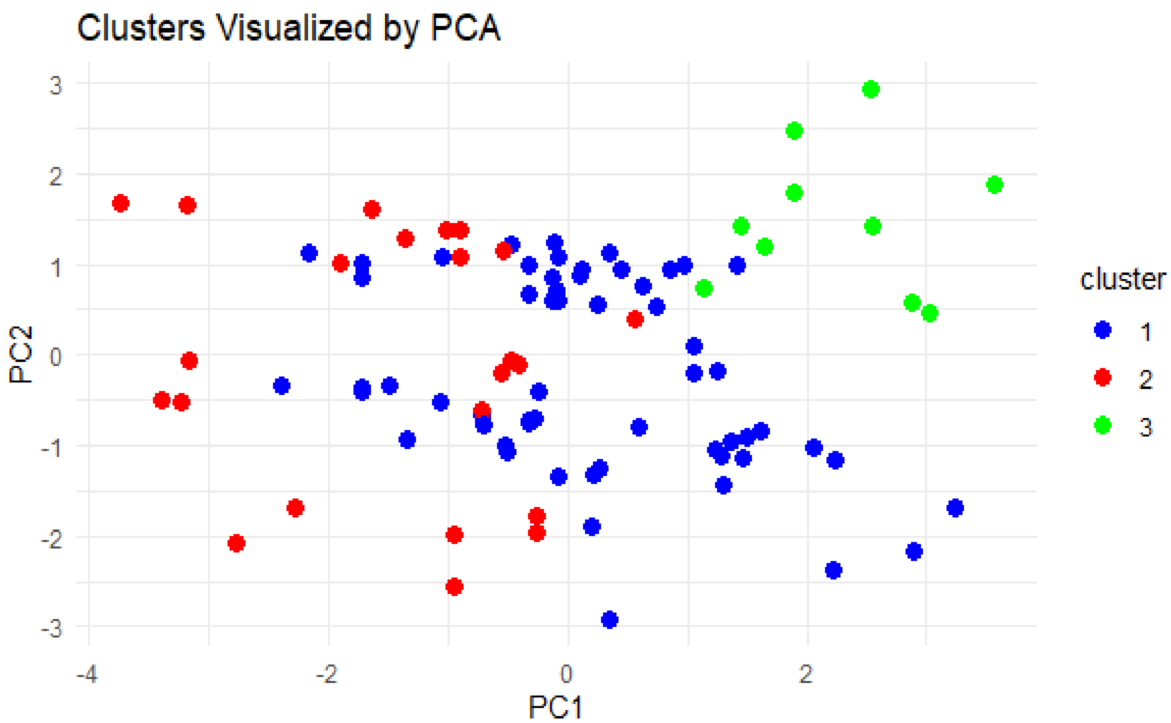
Kmeans clustering

K-means clustering with 3 clusters of sizes 59, 23, 10

Cluster means:

- Cluster TDist SMinutes Calories HeartRate health_score
- 1 6.37288 935.7458 2225.051 76.11864 38.32203
- 2 7.181304 836.6087 3468.391 78.23913 49.47826
- 3 2.026000 448.5000 1218.900 64.60000 29.30000

To visualize the clusters generated by k-means with 5 features, reduce the data to a 2D space for plotting purpose- use **Principal Component Analysis (PCA)**



Interpretation:

- **For** Cluster 1 (Blue)**:**
  - It has individuals leaning towards Cluster 2 or Cluster 3 and they must be provided by personalized recommendations to enhance activity or sustain healthy patterns.
- **For** Cluster 2 (Red)**:**
  - Increased daily steps or reduced sedentary time could improve health outcomes. Higher calorie expenditure through low-intensity activities might be beneficial.
- **For** Cluster 3 (Green)**:**
  - This group is high-performing; they should maintain or enhance their health behaviours for well-being.

Random Forest (Regression)

- Many decision trees form an ensemble model known as random forests.
  random Forest(HealthScore ~ TotalDistance +SedentaryMinutes+ Calories+ HeartRate
- R-squared value of 0.9531

A high R-squared value suggests that the model fits the data very well. In other words, the model is able to explain most of the variation in the target variable

H. Findings

- All the models show that health score dependents on Total distance , Total steps, Calories burnt , Heart Beat Rate and Sedentary minutes .
- These were the features which are majorly used by individuals.
- These models help to predict the health score , when these features are given.
- Users can assess the risk associated with individual with the health data.
- Also groups are formed based on the health score which ensures that high-risk users are correctly identified.

I. Conclusion

Use health scores to assess the risk levels of users. Higher risk scores might indicate a higher likelihood of health problems. Offer personalized health and wellness programs to users based on their health scores. Encourage healthier lifestyles to potentially lower future health problems. By integrating predictive models that calculate health scores, users can enhance their risk assessment capabilities; change their daily habits to improve health outcomes. The proactive approach results in risk management and better user health with potential reductions of future problems thereby improving overall health outcomes.

Recommendation

Uniform health score or a consistent health score and Stable Health Score i.e. if the health score is stable ±2% within a specified range over a 2-month period is a good sign for health. Variable Health Score-if significant fluctuations in the health score over a 2-month period, Minor variations: ±10%, should be considered for improvement and Major Variations: greater than ±10%, should consult the doctor for their health decisions.

ACKNOWLEDGMENT

I would like to express my gratitude to the users who use wearable devices their suggestions and views , that made this research possible. I also want to thank the doctors for his valuable insights and collaborative efforts throughout this project. Finally, I am deeply thankful for the unwavering support of my family and friends.

REFERENCES

1. Willingness to adopt wearable devices with behavioral and economic incentives by health insurance wellness programs: results of a US cross-sectional survey with multiple consumer health vignettes Diego Soliño-Fernandez1 , Alexander Ding2 , Esteban Bayro-Kaiser3* and Eric L. Ding4,5,6,7- * Correspondence: esteban@uni-bremen.de; eding@post.harvard.edu 3 Institute for Artificial Intelligence, University of Bremen, Bremen, Germany 4 Department of Nutrition, Harvard Chan Sc-

hool of Public Health, Boston, MA 02115, USA
2. Wearable data analysis, visualisation and recommendations on the go using android middleware Marios C. Angelides 1 & Lissette Andrea Cabello Wilson1 & Paola Liliana Burneo Echeverría1 - Published : 4 June 2018 , Marios C. Angelides marios.angelides@brunel.ac.uk , Lissette Andrea Cabello Wilson lcabello@fiec.espol.edu.ec 1 Department of Electronic and Computer Engineering, College of Engineering Design and Physical Sciences, Brunel University London, Uxbridge UB8 3PH, UK
3. Using Wearables and Machine Learning to Enable Personalized Lifestyle Recommendations to Improve Blood Pressure.- Chiang, Po-Han, Melissa Wong, and Sujit Dey. IEEE Journal of Translational Engineering in Health and Medicine 9 (2021): 1–13
4. Cognitive Training and Stress Detection in MCI Frail Older People Through Wearable Sensors and Machine Learning.- Delmastro, Franca, Flavio Di Martino, and Cristina Dolciotti. IEEE Access 8 (2020): 65573–90.
5. A Secure AI-Driven Architecture for Automated Insurance Systems: Fraud Detection and Risk Measurement NAJMEDDINE DHIEB1 , (Student Member, IEEE), HAKIM GHAZZAI 1 , (Senior Member, IEEE), HICHEM BESBES 2 , AND YEHIA MASSOUD1 , (Fellow, IEEE) 1School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ 07030, USA 2Higher School of Communications of Tunis, University of Carthage, Ariana 2083, Tunisia Corresponding author: Hakim Ghazzai (hghazzai@stevens.edu)
6. HEALTH RISK PREDICTION BY MACHINE LEARNING OVER DATA ANALYTICS PRABHU. T1, DARSHANA. J 2, DHARANI KUMAR. M3, HANSAA NAZREEN. M4 1Assistant Professor Department of ECE, Department of Electronics and Communication Engineering, SNS College of Technology, Coimbatore, Tamil Nadu, India 2, 3 4Final Year ECE Students, Department of Electronics and Communication Engineering, SNS College of Technology, Coimbatore, Tamil Nadu, India
7. Machine Learning for Healthcare Wearable Devices: The Big Picture.- Sabry, Farida, Tamer Eltaras, Wadha Labda, Khawla Alzoubi, and Qutaibah Malluhi. Journal of Healthcare Engineering 2022 (April 18, 2022): e4653923.
8. A Survey of Wearable Devices and Challenges- Suranga SeneviratneThe University of Sydney, Tham Nguyen- University of Technology Sydney Guohao Lan- Delft University of Technology,Article in IEEE Communications Surveys and Tutorials -July 2017
9. Shin, Grace, Mohammad Hossein Jarrahi, Yu Fei, Amir Karami, Nicci Gafinowitz, Ahjung Byun, and Xiaopeng Lu. "Wearable Activity Trackers, Accuracy, Adoption, Acceptance and Health Impact: A Systematic Literature Review." Journal of Biomedical Informatics 93 (May 1, 2019): 103153.
10. Wearable data analysis, visualization and recommendations on the go using android middleware - Marios C. Angelides , Lissette Andrea Cabello Wilson,Paola Liliana Burneo Echeverr´ıa1, Department of Electronic and Computer Engineering, College of Engineering Design and Physical Sciences, Brunel University London, Uxbridge UB8 3PH, UK - Springer Journal-2018.
11. Survey on: Applications of Smart Wearable Technology in Health Insurance Apeksha Shah1, Dr. Swati Ahirrao2, Dr. Shraddha Phansalkar2, Dr. Ketan Kotecha3 Symbiosis Institute of Technology, Symbiosis International (Deemed University)-Pune-2021 IOP Conference Series: Materials Science and Engineerin