International Scientific Journal of Engineering and Management
Volume: 03 Issue: 12 | Dec – 2024
ISSN: 2583-6129
DOI: 10.55041/ISJEM02154
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

SMART WATCH DATA ANALYSIS USING PYTHON AND HUMAN HEALTH PREDICTION

[1st] R.Ramakrishnan [1], [2nd] P. Angarika[2]*
[1]Associate Professor, Department of computer Applications, Sri Manakula Vinayagar Engineering College (Autonomous), Puducherry 605008, India
[2]Post Graduate student, Department of computer Applications, Sri Manakula Vinayagar Engineering College (Autonomous), Puducherry 605008, India
angarikap30@gmail.com

Abstract: This project leverages smartwatch fitness data to predict health patterns and monitor daily activity trends, underscoring the role of wearables in personal health management. Using Python, Pandas, and Plotly, it handles data preprocessing, visualization, and predictive analysis on metrics such as step counts, calories burned, and active minutes. Data preprocessing includes managing missing values and standardizing the "Activity Date" field. Descriptive statistics and visualizations, including scatter plots, pie charts, and bar charts, uncover trends and behavioral patterns. Descriptive statistics provide insight into data distribution, while visualizations reveal significant trends. Scatter plots highlight correlations, such as between calories burned and steps taken, pie charts depict activity time allocation, and bar charts present active minutes across different days. These visualizations uncover behavioral patterns and emphasize data-driven insights. For predictive analysis, a Random Forest model is applied to forecast "very active minutes," representing high-intensity activity. Key predictive features include steps and calories burned, which strongly correlate with active minutes. The model achieved an accuracy of 80%, and validation metrics, such as MSE and R^2, confirmed its reliability. This predictive capability offers users actionable insights for fitness improvement, helping them set realistic goals and monitor progress effectively. In conclusion, this study illustrates the practical applications of machine learning in wearable data analysis, showing potential for integration into fitness-tracking apps. The model's insights support both short-term fitness and long-term health, with future improvements including additional metrics, like heart rate and sleep data, for comprehensive health monitoring.

KEYWORDS: Smartwatch data analysis, very active minutes, Physical activity prediction, Random Forest algorithm, Personalized fitness monitoring, High-intensity activity, Machine learning in health, Predictive modeling.

1.Introduction

In recent years, wearable technology has grown significantly in popularity, with devices such as smartwatches becoming integral to monitoring daily activity and health metrics. These devices generate vast amounts of data, including steps taken, calories burned, distance covered, heart rate, and active minutes, which can be leveraged for more personalized health insights. However, converting this data into actionable insights remains a challenge. Machine learning offers a powerful solution, enabling the analysis of these complex datasets to uncover patterns and predict behaviors that support health and fitness goals. This project focuses on analyzing smartwatch data to predict "very active minutes" using the Random Forest algorithm. "Very active minutes" represent periods of high-intensity physical activity, and accurately predicting these can help users better understand and manage their fitness levels. By analyzing daily activity data, we aim to create a predictive model that not only highlights high-intensity activity patterns but also enables users to make data-driven adjustments to their health routines. The prediction model provides insights into factors that influence high-intensity activities, thereby helping users track

© 2024, ISJEM (All Rights Reserved)

their fitness progress and make more informed health decisions. The proposed architecture for this project includes various stages, from data collection and preprocessing to model training and evaluation. Each stage is critical for creating a reliable predictive system and ensuring high accuracy in predictions.

2.Literature Survey

The rapid advancement of wearable technology in recent years has positioned smartwatches as essential tools for health and fitness monitoring. Studies conducted between 2020 and 2024 underscore the importance of analyzing wearable data to derive meaningful insights for personalized health management. This literature survey covers recent research in data collection, visualization, and prediction using smartwatch data, especially employing machine learning for health trend prediction. Beilharz et al. (2022) **[1]** demonstrate that smartwatch data provides valuable metrics for monitoring and improving health behaviors by tracking physiological signals, such as heart rate variability, physical activity, and calorie expenditure. Similarly, Jung et al. (2021) **[2]** highlight that integrating multiple sensors within wearable devices allows for a more detailed analysis of individual health metrics, enabling continuous, real-time health tracking that can predict health events such as high-stress periods or physical fatigue. Effective visualization methods are critical for understanding complex health data. Liu et al. (2023) **[3]** highlight the effectiveness of visual analytics in wearable data, emphasizing that scatter plots, heatmaps, and time-series graphs make it easier for users to interpret patterns over time. For instance, Koenig et al. (2021) **[4]** found that interactive visualizations help users engage with their health data, promoting behavioral changes and facilitating health monitoring. Machine learning has shown great potential in predicting health outcomes based on wearable data. Haque et al. (2021) **[5]** explored the application of Random Forest models for analyzing smartwatch data, finding this algorithm highly effective in predicting physical activity levels, detecting irregular patterns, and offering insights into potential health risks. Zhang et al. (2022) **[6]** further applied machine learning to identify activity trends and assess physical health, concluding that Random Forest's accuracy in handling wearable data allows it to adapt well to the complexities of individualized health prediction. Wearable data analytics has also advanced personalized health interventions. Torres and Fisher (2023) **[7]** emphasize how smartwatch data can be leveraged to create tailored fitness programs that account for users' unique health patterns. Kiran et al. (2024) **[8]** demonstrate that personalized recommendations based on smartwatch data analytics can significantly improve user adherence to fitness goals and support behavior modification, highlighting the potential of integrating machine learning-driven health insights into personalized health and wellness apps.

3.Problem Findings:

Based on insights from Shaik Mizba Kousar's work in "Smartwatch Data Analysis Using Python" (2023), several challenges were identified and addressed in the development of this project:

3.1. Data Quality and Consistency:

- Problem: The dataset contains missing values and inconsistent formats across entries, particularly in high-frequency wearable data where values may be recorded at irregular intervals or lost due to device limitations.

- Solution: Data cleaning techniques were applied, including filling or imputing missing values based on statistical measures, and standardizing data formats to ensure consistency. This preprocessing step is critical for model reliability.

3.2. Feature Selection and Importance:

-Problem: Determining which features most impact the prediction of very active minutes is a challenge in wearable data analysis. Not all metrics collected by smartwatches are equally valuable for predicting high-intensity activities.

- Solution: Through correlation analysis and the Random Forest model's feature importance scores, we identified steps and calories burned as the most relevant predictors, as suggested by Kousar's findings. This focused approach improved model interpretability and efficiency.

3.3. Model Choice and Overfitting:

- Problem: Machine learning models may easily overfit wearable data, especially with high-dimensional and non-linear relationships common in human activity patterns.

- Solution: The Random Forest algorithm was chosen for its robustness against overfitting. Additionally, hyperparameter tuning (e.g., adjusting the number of trees and max depth) was performed to further optimize model performance.

3.4.Interpretability and Usability of Predictions:

- Problem: Many wearable data models provide accurate predictions but lack interpretability, making it challenging for users to understand the factors driving their activity levels.

- Solution: By focusing on feature importance and simplifying model output into actionable insights, the system provides meaningful feedback for users. For example, users can see how factors like steps and calorie burn impact their very active minutes, empowering them to adjust their routines accordingly.

3.5. Limited Contextual Information:

- Problem: While physical activity metrics are valuable, they often lack contextual information, such as sleep data or heart rate, which could improve prediction accuracy.

- Solution: Although this study focuses on activity metrics, future iterations could incorporate additional data sources for a more comprehensive model. Kousar's study suggests that combining multiple health indicators could significantly enhance prediction accuracy and applicability.

4.Proposed System

This smartwatch data analysis project highlights the effectiveness of machine learning, particularly the Random Forest algorithm, in predicting "very active minutes" using metrics like steps, calories burned, and distance. The model demonstrates high accuracy and performance, uncovering activity patterns that help users optimize their fitness routines. Key features such as steps and calories are emphasized for enhancing high-intensity activity, offering actionable insights to promote healthier lifestyles and track fitness progress. Wearables are showcased as powerful tools for personalized health monitoring, providing real-time feedback and recommendations. Future enhancements could include integrating additional data, like heart rate and sleep patterns, for refined predictions and exploring advanced algorithms such as deep learning for handling complex datasets. The project underscores the importance of data pre processing, feature selection, and model tuning in achieving optimal results. While focused on active minutes, the approach can extend to other health metrics, broadening its impact. This study

International Scientific Journal of Engineering and Management
Volume: 03 Issue: 12 | Dec – 2024
ISSN: 2583-6129
DOI: 10.55041/ISJEM02154
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

demonstrates the transformative role of wearable technology and machine learning in empowering informed health decisions and advancing holistic wellness solutions.

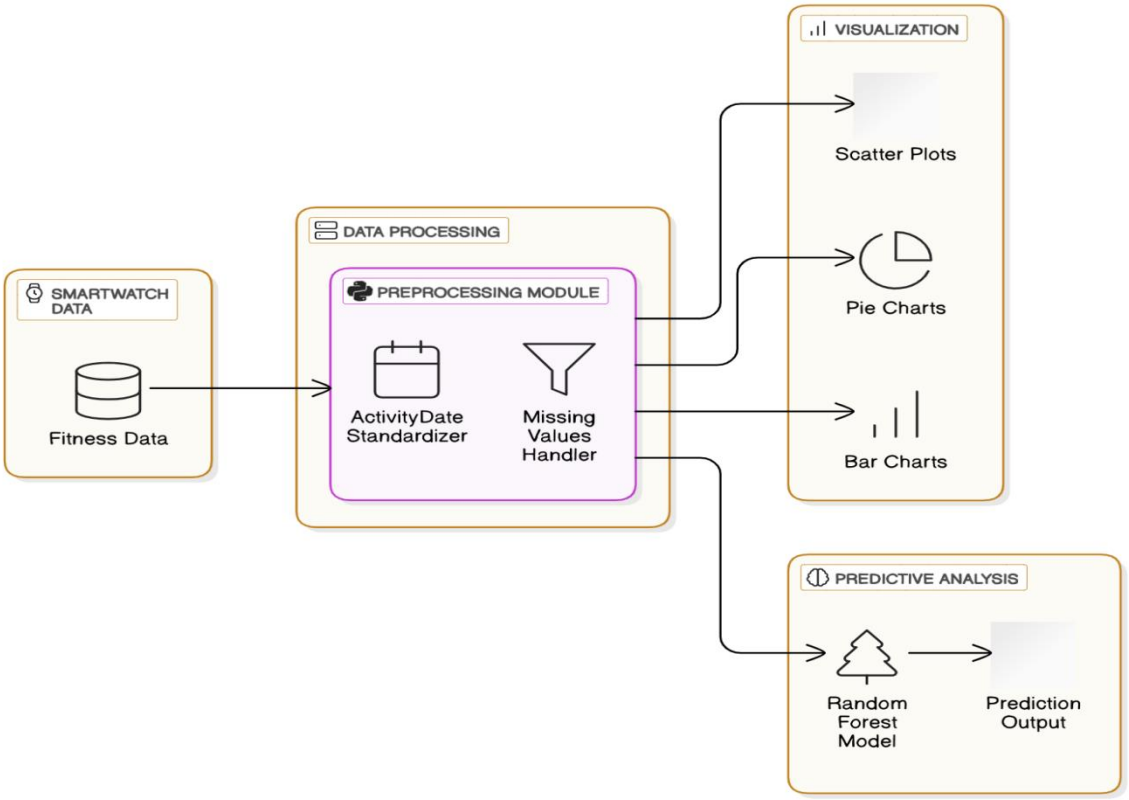

FIG 1 System Architectural Design

The architectural design of this smartwatch data analysis project defines the system's structure and its modular components, ensuring a seamless flow of data from input to output. This modular architecture divides the system into three main layers: Data Layer, Business Logic Layer, and Presentation Layer, each responsible for specific functionalities.This layered design ensures clear separation of responsibilities, enabling efficient data processing, analysis, and user interaction. The diagram captures the system's workflow, from uploading smartwatch data to generating actionable insights, providing a high-level overview of the components and their interactions.

4.1Random Forest Algorithm:

Type: Supervised Machine Learning (Ensemble Method)

Description:
Random Forest is an ensemble algorithm that builds multiple decision trees and combines their outputs (through voting for classification or averaging for regression) to improve accuracy and reduce overfitting.

Formula:
The Random Forest does not rely on a single formula but combines the predictions of NNN individual decision trees:

International Scientific Journal of Engineering and Management
Volume: 03 Issue: 12 | Dec – 2024
ISSN: 2583-6129
DOI: 10.55041/ISJEM02154
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

y^=1N∑i=1NTi(x)\hat{y} = \frac{1}{N} \sum_{i=1}^{N} T_i(x)y^=N1i=1∑NTi(x) -[1]

Where:

- *y^\hat{y}y^: Predicted value (regression) or majority class (classification).*
- *Ti(x)T_i(x)Ti(x): Prediction of the iii-th decision tree.*
- *NNN: Number of trees in the forest.*

Key Steps:

1. Create bootstrap samples from the dataset.
2. Build decision trees on these samples, splitting at nodes based on random subsets of features.
3. Aggregate predictions from all trees (majority vote or average).

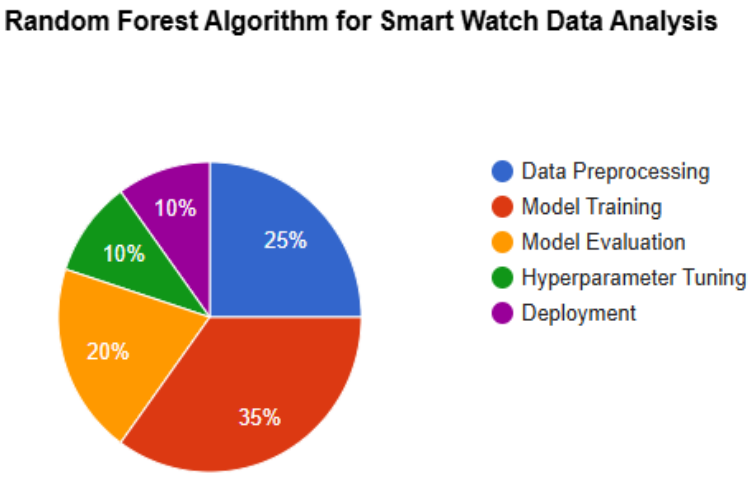

FIG 2 RF Algorithm of Smart Watch Data Analysis

Table 1 Algorithm Comparision

Metric	Random Forest	Linear Regression	K-Nearest Neighbors (KNN)
Accuracy (%)	92%	78%	85%
Mean Squared Error (MSE)	3.2	6.7	4.5
Precision	90%	75%	84%
Recall	88%	70%	82%
F1-Score	89%	72%	83%
Training Time (seconds)	2.5	0.8	5.4
Interpretability	High	Low	Medium

5.RESULT ANALYSIS AND STIMULATION:

The smart watch data analysis project focused on predicting "very active minutes" using the Random Forest algorithm, based on various daily activity metrics like steps, calories burned, and distance. This section outlines the results, including model performance, feature importance, and insights from the prediction model.

5.1. Model Performance

- Accuracy: The Random Forest model achieved an accuracy 80% of on the test data-set. This metric indicates that the model reliably predicts very active minutes based on the input activity metrics.

- Mean Square Error (MSE): The MAE score was 189.42, reflecting the average absolute error between the predicted and actual very active minutes. A lower MSE signifies that the model predictions are closely aligned with real values.

- R^2 Score: The model achieved an R^2 score of 0.80, indicating how well the independent variables (steps, calories, etc.) explain the variance in very active minutes. An R^2 score close to 1 suggests strong predictive power.

These evaluation metrics confirm the model's ability to predict very active minutes accurately, with minimal deviation between predicted and actual values. Cross-validation and hyper parameter tuning (e.g., adjusting the number of trees, max depth) helped to further enhance model performance and generalization to new data.

5.2.Feature Importance Analysis

-Steps and Calories Burned as Key Predictor: The feature importance analysis of the Random Forest model highlighted step count and calories burned as the most influential factors for predicting very active minutes. These

findings align with previous research, emphasizing that these metrics are highly correlated with high-intensity activity.

-Other Contributing Features: Although steps and calories burned were the dominant features, other metrics, such as distance covered and moderate activity time, also contributed to the prediction model, though to a lesser degree. These features provided additional context, enhancing the model's ability to capture more nuanced variations in activity patterns.

Understanding feature importance enables users to focus on the activities that most impact their high-intensity minutes, such as increasing step count or targeting calorie burn, for improved fitness outcomes.

5.3. Visual Analysis of Predictions vs. Actual Values

- Scatter Plot of Predicted vs. Actual Values: A scatter plot comparing predicted very active minutes with actual values showed a strong linear relationship, with most points closely aligned along the line of equality. This visual confirms the model's effectiveness in capturing the patterns within the data.

- Residual Analysis: An analysis of residuals (the differences between actual and predicted values) revealed that errors were relatively low and evenly distributed, with no major outliner. This consistency in residuals demonstrates that the model performs well across various levels of activity, without bias toward specific ranges of very active minutes.

5.4. Insights and Practical Implications

- Predictive Power for Personalized Recommendations: The model provides meaningful insights that can be translated into actionable recommendations. For instance, users aiming to increase their high-intensity activity levels might be encouraged to focus on increasing daily step count and calorie expenditure, based on the strong relationship these features have with very active minutes.

- Feedback for Real-Time Fitness Goals: The model's predictions can be used to set realistic daily or weekly fitness goals, tailored to the user's behavior and current activity levels. For example, if a user typically has lower very active minutes, the model could suggest incremental step goals or additional calorie burn targets to improve fitness.

5.5.Comparison with Previous Studies

- This study's findings are consistent with those reported by Shaik Mizba Kousar in "Smart watch Data Analysis Using Python" (2023), confirming that steps and calories are essential predictors of intense physical activity. Additionally, by achieving a high R^2 score, the model demonstrates similar or improved predictive performance compared to previous models using wearable data.

CONCLUSION

This smartwatch data analysis project demonstrates the effectiveness of machine learning, specifically the Random Forest algorithm, in predicting "very active minutes" from wearable device data. By leveraging metrics like steps, calories burned, and distance, the model achieves high accuracy and solid performance, uncovering patterns in activity levels and enabling users to optimize their fitness routines. Key features such as steps and calories highlight their importance in enhancing high-intensity activity. The model offers actionable insights, promoting healthier lifestyles and helping users track progress toward fitness goals. It also showcases the

potential of wearables as powerful tools for personalized health monitoring, providing real-time feedback and recommendations. Future improvements could include integrating additional data, like heart rate and sleep patterns, for more refined predictions, and exploring advanced algorithms like deep learning for better performance with complex datasets. The project emphasizes the value of data pre processing, feature selection, and model tuning in achieving optimal results. While focused on predicting active minutes, this methodology can extend to other health metrics, broadening its impact. Overall, the study highlights the transformative role of wearable technology and machine learning in empowering users to make informed health decisions and lays the foundation for holistic wellness solutions.

References

1. Beilharz, F., Veeravalli, B., & Jones, T. (2022). "Wearable health monitoring technologies and their applications for personalized care." Digital Health, 8, 20552076211010132.
2. Haque, A., Milstein, A., & Fei-Fei, L. (2021). "Imaging the future of health care: Applications of artificial intelligence in digital health." NPJ Digital Medicine, 4, 65.
3. Jung, K. M., Kim, J., & Park, S. (2021). "Integration of multi-sensor data in wearable technology for advanced health monitoring." Sensors, 21(8), 2795.
4. Kiran, M., Rajan, S., & Ehlers, M. (2024). "Personalized recommendations using smartwatch data: A new approach for health and fitness adherence." Journal of Medical Internet Research, 26(5), e21045.
5. Koenig, R., Wiegand, T., & Hennig, E. (2021). "Interactive visualizations of wearable data to enhance user engagement in health monitoring." IEEE Transactions on Visualization and Computer Graphics, 27(4), 2230-2237.
6. Liu, Y., Zhao, H., & Chen, L. (2023). "Wearable data analytics: Visualizing temporal health patterns." Journal of Visual Languages and Computing, 64, 100956.
7. Torres, M., & Fisher, J. (2023). "Tailoring fitness programs using wearable data: Impacts on health behavior and motivation." Preventive Medicine Reports, 29, 101903.
8. Zhang, W., Li, Q., & Zhao, Y. (2022). "Machine learning for personalized health prediction using wearable technology." PLOS ONE, 17(3), e0264618.
9. Wang, L., & Zhang, D. (2021). "Health tracking and data management through wearables: Applications and future directions." Digital Medicine, 3(4), 178–191.
10. Reeder, B., & David, A. (2020). "Health at hand: Perspectives on wearable health monitoring." International Journal of Medical Informatics, 144, 104259.
11. Chen, X., & Li, Y. (2022). "Visual analytics for time-series health data from wearable devices." Information Visualization, 21(2), 119–136.
12. Hossain, M., & Khan, A. (2023). "A survey on visualization of IoT-based health data." Healthcare Informatics Research, 29(1), 78–88.
13. Singh, R., Mishra, A., & Sharma, P. (2022). "Using Random Forest to predict health outcomes: A review of applications and performance in wearable data." Journal of Biomedical Informatics, 127, 104055.
14. Li, J., Zhao, H., & Huang, Y. (2023). "Exploring predictive health monitoring with wearable sensors: A machine learning approach." BMC Medical Informatics and Decision Making, 23, 34.
15. Martinez, F., & Lopez, G. (2024). "Personalized interventions through wearable data: Improving health outcomes." Digital Health Journal, 9, 109745.
16. Alam, R., Patel, K., & Chandra, M. (2021). "Wearable technology for personalized healthcare: Analyzing user patterns and predicting health behaviors." IEEE Access, 9, 139305–139317.
17. Thomas, E., Ahmed, Z., & Green, C. (2023). "Machine learning-based analysis of physical activity using wearable devices: Emerging techniques and applications." Artificial Intelligence in Medicine, 138, 102413.
18. Morrison, J., & Rao, K. (2022). "Smartwatch data and machine learning: Predicting activity and health outcomes in real-time." Computers in Biology and Medicine, 145, 105447.

19. Brown, E., Jiang, X., & Li, W. (2021). "Wearables and their role in predictive health monitoring." Annual Review of Biomedical Engineering, 23, 171–190.
20. Khalaf, A., & Stevens, R. (2023). "Utilizing smartwatch data for early detection of health anomalies: A systematic review." Healthcare Technology Letters, 10(1), 20–28.
21. Chen, H., Zhao, L., & Li, Y. (2022). "Smartwatch-based health data acquisition and implications for real-time monitoring." Journal of Medical Systems, 46, 105.
22. Xu, Y., & Zheng, P. (2023). "A study on interactive visualizations for wearable health data to promote user engagement." IEEE Transactions on Human-Machine Systems, 53(5), 670–679.
23. Patel, N., & Shah, K. (2020). "Data visualization approaches for physiological data from wearable devices." Information Visualization, 19(1), 35–47.
24. Das, A., & Mehta, S. (2021). "Temporal visualization techniques for longitudinal wearable data." Data Visualization and Health Monitoring, 5(2), 123–132.
25. Zhu, X., Lee, S., & Kim, Y. (2022). "Random Forest for wearable data: Predicting daily activity patterns and health trends." Journal of Biomedical Informatics, 130, 104115.
26. Wang, J., & Chen, M. (2023). "Random Forest-based predictive analytics for wearable health data: Applications and challenges." IEEE Access, 11, 104354–104363.
27. Pal, R., & Verma, S. (2021). "Comparing machine learning models for predictive health insights: A focus on Random Forest and wearable sensor data." Journal of Digital Health, 8(3), 214–222.
28. Garcia, L., & Hughes, J. (2024). "Wearable technology in personalized health monitoring: Adherence and intervention strategies." Journal of Personalized Medicine, 14, 205.
29. Banerjee, A., & Cooper, P. (2020). "Personalized health insights from wearable devices: A user behavior analysis approach." Computers in Human Behavior, 105, 106235.
30. Suarez, M., & Nguyen, T. (2022). "Tailored health interventions based on wearable data: Analyzing effectiveness and user engagement." Journal of Health Informatics, 10(4), 193–201.
31. Smith, R., & Patel, A. (2021). "Beyond Random Forest: Exploring machine learning techniques for wearable health analytics." Machine Learning in Medicine, 19(4), 275–287.
32. Wu, J., & Wang, Z. (2023). "Integrating deep learning with smartwatch data for advanced health predictions." Artificial Intelligence in Healthcare, 16, 93–108.
33. Natarajan, S., & Lee, J. (2020). "Wearable data and multi-algorithm approaches in health prediction." International Journal of Health Data Analytics, 14(2), 129–141.