Human Activity Recognition

"Unlocking Your Life's Data: Transforming Raw Sensor Insights into Personalized Health Monitoring and Adaptive Technology Experiences."

Background/Motivation

Human Activity Recognition (HAR) using accelerometer data is a pivotal research area with applications in health monitoring, sports optimization, and context-aware computing. This endeavor holds significance due to its wide-ranging applications. The resulting HAR system could be pivotal in health and wellness monitoring, sports optimization, and context-aware computing. By automatically identifying and categorizing human activities, it facilitates personalized fitness plans, early detection of health issues, and enhanced user experiences in smart environments. Additionally, in fields like sports, healthcare, and assistive technologies, the ability to recognize and respond to human activities can lead to more efficient training, improved safety, and better support for individuals with mobility impairments.

Problem Formulation

The task at hand involves constructing a Human Activity Recognition (HAR) system using a dataset that includes 3-axis accelerometer data from four distinct body locations: waist, left thigh, right arm, and right ankle. The objective is to assess the feasibility of creating a wearable device comparable to those offered by BodyMedia and Fitbit.

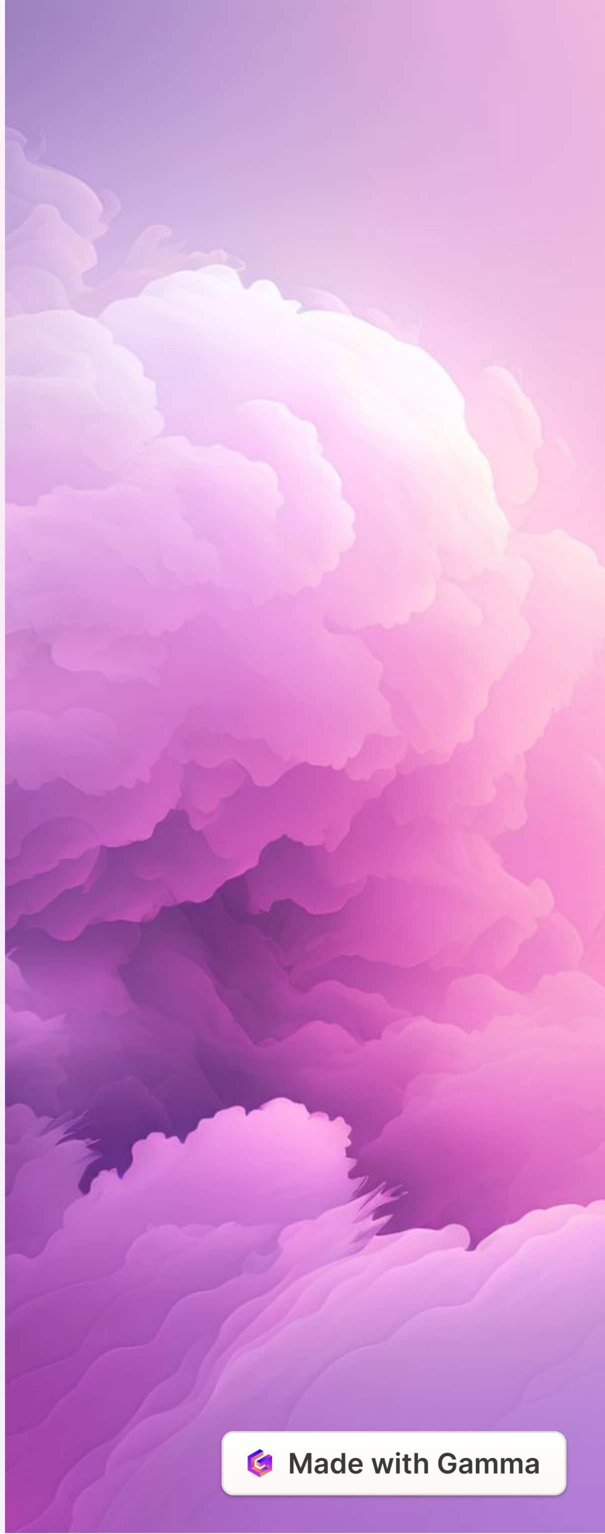
The problem at hand is framed as a supervised machine learning classification task. The objective is to map accelerometer data from various body locations (waist, left thigh, right arm, and right ankle) to specific human activities, categorized into five classes: sitting, sitting-down, standing, standing-up, and walking. This falls within the realm of supervised learning, as the model is trained on labeled examples to predict the activity class for unseen instances.

The nature of the problem is a classification task, as the goal is to categorize the activities into predefined classes. This involves training a machine learning model on the provided dataset to learn the patterns and relationships between accelerometer features and the corresponding human activities. The success criteria for the model will be its ability to accurately predict the activity class based on new accelerometer data, thus aiding in the construction of a wearable device for real-time activity monitoring.

Training Data Description:

- **Unit of Analysis:**
  - Each row in the training data table represents a time window of 150ms, capturing sequential 3-axis accelerometer measurements from one of four body locations (waist, left thigh, right arm, or right ankle).
- **Input Variables:**
  - The input variables consist of the 3-axis acceleration measurements (X, Y, Z) recorded by the accelerometers for each body location.
- **Output/Labels:**
  - The output or labels correspond to the activity class performed during the time window. Each row is labeled with the specific human activity, such as sitting, sitting-down, standing, standing-up, or walking.

In summary, each row in the training data represents a discrete time interval, with the input variables comprising 3-axis accelerometer measurements, and the output being the corresponding human activity label for that time window.



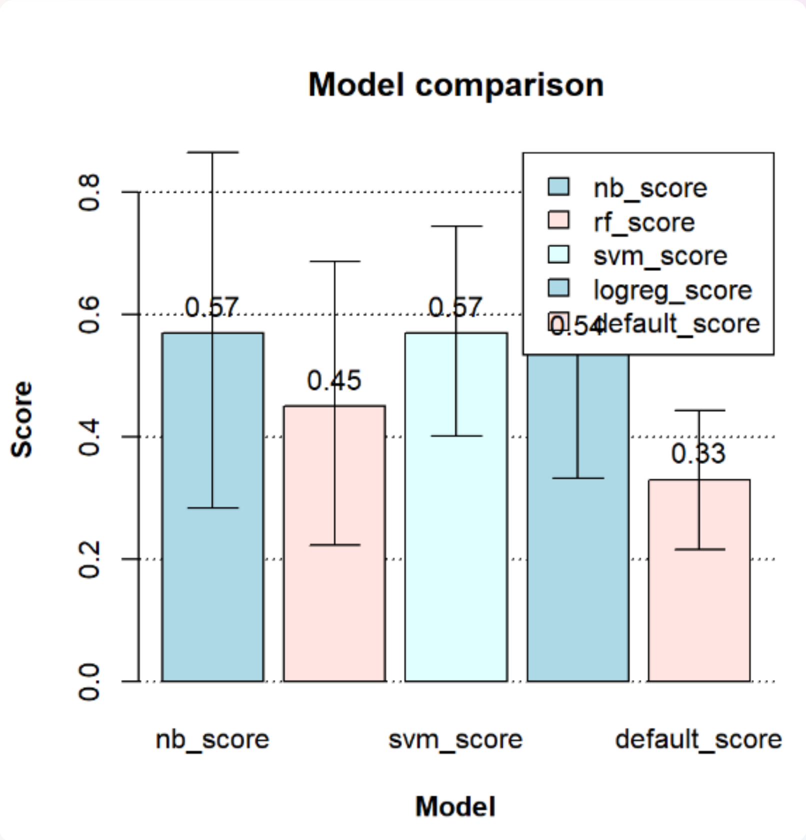
Data preprocessing

Data preprocessing is a crucial step in preparing raw datasets for analysis or modeling. In the provided code snippet, the focus is on preprocessing the 'z4' column within a DataFrame. The following steps are undertaken:

1. **Conversion to Numeric:**
   - The 'z4' column is converted to a numeric format using `pd.to_numeric()`.
   - The `errors='coerce'` parameter is utilized to handle non-numeric values by replacing them with NaN.
2. **Handling Missing Values:**
   - Any remaining NaN values in the 'z4' column are filled with the mean of the column using `fillna()`.

These preprocessing steps are essential for ensuring data integrity and model compatibility. By converting non-numeric values and handling missing data appropriately, the 'z4' column becomes suitable for downstream analysis or machine learning tasks. The use of mean imputation for missing values is a common strategy, providing a reasonable estimate while maintaining the overall statistical characteristics of the data.

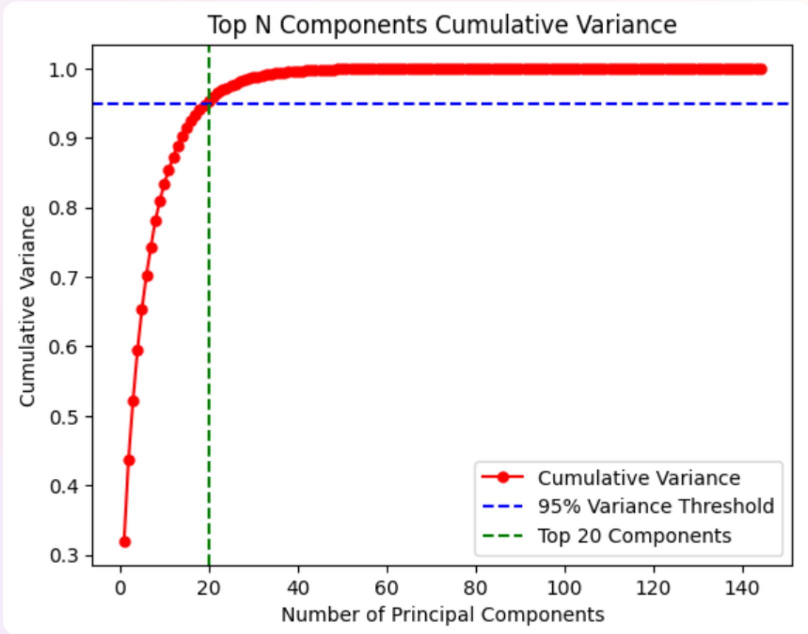
Initial ML analysis



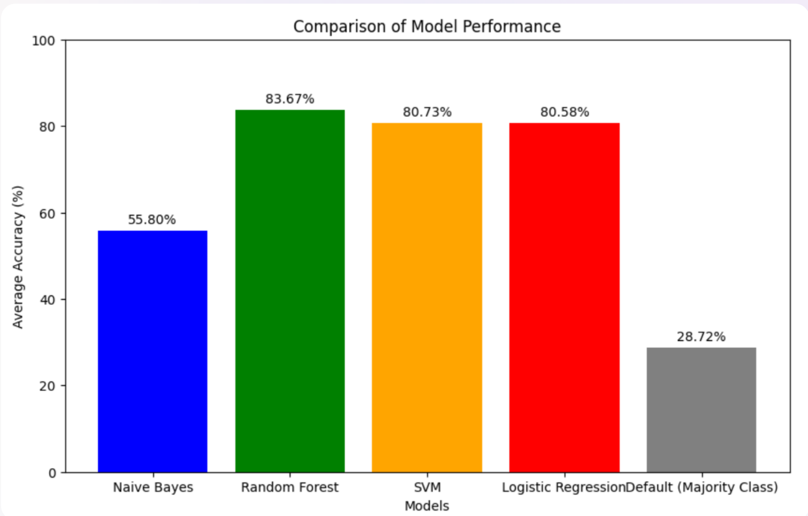
Developed different machine learning models initially but the performance of different models were not satisfactory hence additional analysis was required with new features.

Model Development

Performed PCA analysis and found that 95% of the variance is explained by the top 20 components which can be further used for the analysis.



ML model comparison



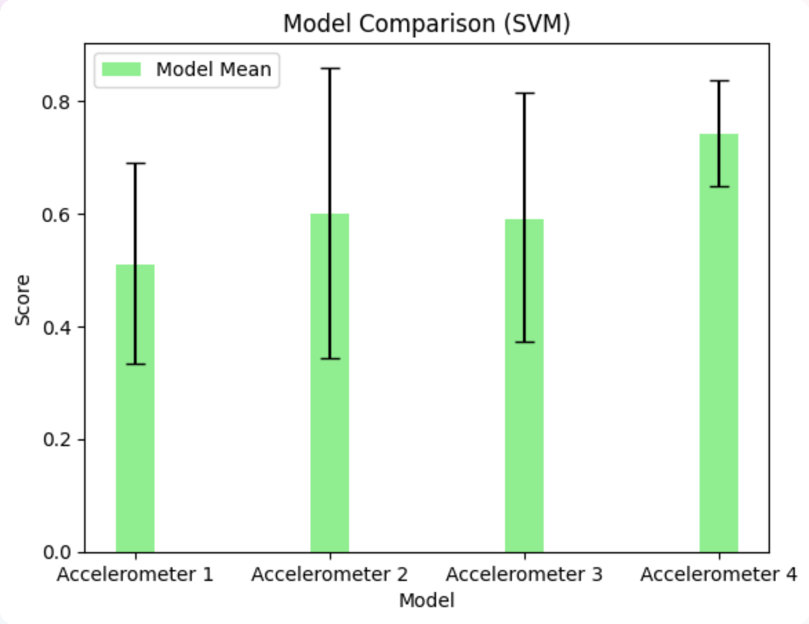
4 fold cross-validation framework is implemented for Human Activity Recognition (HAR). It preprocesses accelerometer data, applies PCA for dimensionality reduction, and evaluates the performance of various classifiers. Results show that Random Forest achieves the highest average accuracy at 83.7%, while Naive Bayes performs less effectively with an average accuracy of 55.8%. Support Vector Machine and Logistic Regression exhibit competitive average accuracies of 80.7% and 80.6%, respectively. The baseline accuracy, obtained by predicting the majority class, is 28.7%. These findings inform the selection and optimization of machine learning models for HAR applications.

Accelerometer Model Comparison

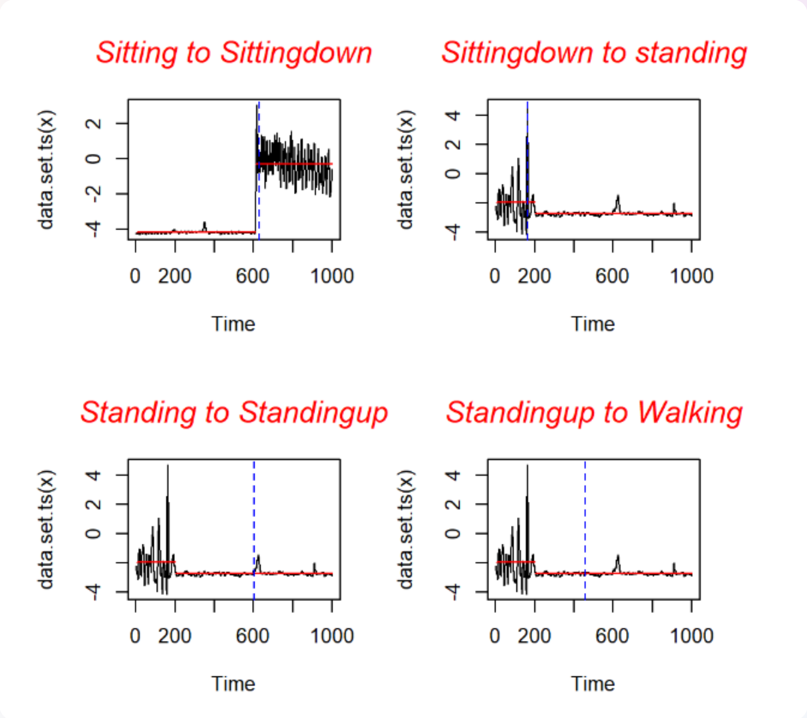
Accelerometer 4(right ankle) Outperforms Others: SVM scores for Accelerometer 4 exhibit the highest mean accuracy among all locations.

Average Accuracy for Accelerometer 4: 74.24%.

Varied Performance Across Locations:Accelerometer 2 and Accelerometer 3 show competitive results, with mean accuracies of 60.16% and 59.41%, respectively.



Transition from one activity to another

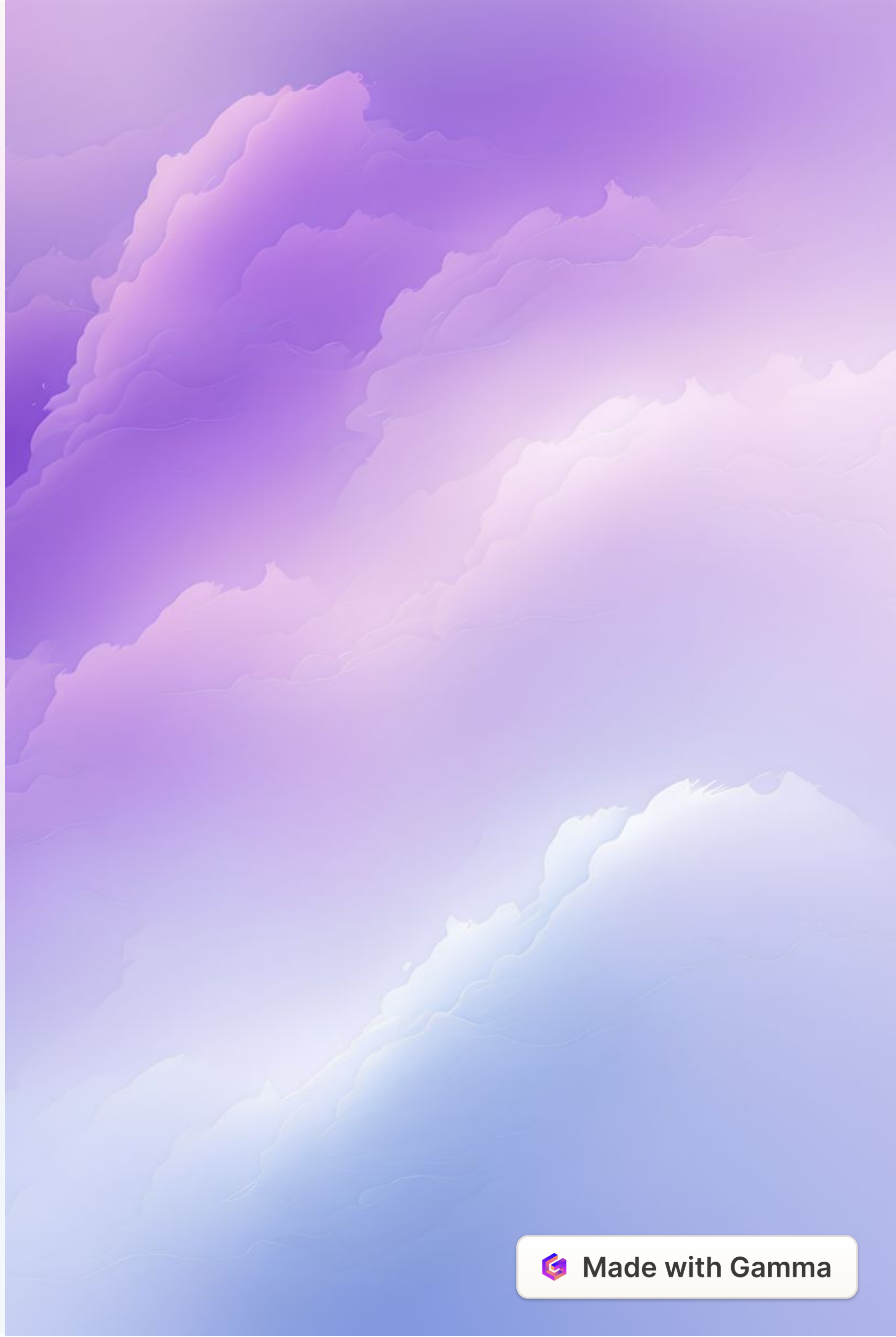


In change of detection ,sitting down and standing can be detected fairly quickly, whereas, standing up and walking may take up to 15s to detect.

Made with Gamma

Results and Insights

Our data samples only contains four individuals activities across 8 hours. Insufficient data point and user sample may lead into a biased conclusion.. The raw features along couldn’t not provide a accurate model, instead, feature Engineering brought significant improvement on classification accuracy over raw features, so I consider this as a effective approach. The final classifier with feature engineering can be used to predict activities very well. Meanwhile, I also found the best positions to have accelerator are right ankle and left thigh, which provide equal accuracy when using along. In terms of change point detection, sitting down and standing can be detected fairly quickly, whereas, standing up and walking may take up to 15s to detect. The takeaway from this project is that I was able to build a foundation of validation, features engineering, change point detection with a couple of prediction models such as Random Forest, Multinomial SVM, Multinomial Logistic Regression as well as Naïve Bayes.



Conclusion

In summary, the Human Activity Recognition (HAR) project aimed to construct an accurate model for classifying activities from accelerometer data. By exploring temporal features and change-point detection, the system's responsiveness was enhanced. The future holds potential for advanced applications, including personalized models, multi-sensor integration, and real-time implementation. As technology progresses, HAR's impact is poised to extend across rehabilitation, smart homes, occupational safety, virtual reality, and various other domains. Through this ongoing evolution, HAR stands to revolutionize how we monitor and interact with human activities in diverse and meaningful ways.

Future applications

1. **Rehabilitation Monitoring:**
    - Extend the use of HAR technology to monitor and guide rehabilitation exercises, providing real-time feedback to individuals recovering from injuries or surgeries.
2. **Aging-in-Place Solutions:**
    - Implement HAR in smart homes to support aging-in-place by autonomously detecting changes in daily activities and alerting caregivers or healthcare providers to potential health issues.
3. **Occupational Safety:**
    - Apply HAR in industrial settings to enhance occupational safety by monitoring workers' activities and providing immediate alerts for potential hazards or fatigue.
4. **Virtual Reality Interaction:**
    - Integrate HAR into virtual reality systems for more natural and immersive user interactions, enabling gestures and movements to control virtual environments.
5. **Educational Environments:**
    - Utilize HAR in educational settings to track students' physical engagement and activities, providing insights for adaptive teaching strategies and promoting healthier learning environments.

YouTube video

Watch Now

References

1. *Ronao, C. A., & Cho, S. (2016, June 1). Deep learning for human activity recognition: A resource efficient implementation on low-power devices. IEEE Conference Publication | IEEE Xplore.* **https://ieeexplore.ieee.org/document/7516235**
2. *Lara, O. D. (2013). A Survey on Human Activity Recognition using Wearable Sensors.* **https://www.semanticscholar.org/paper/A-Survey-on-Human-Activity-Recognition-using-Lara-Labrador/8d3041129b500b90521c7d768996fc2de11b0e47**
3. *Ronao, C. A., & Cho, S. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems With Applications, 59, 235–244.* **https://doi.org/10.1016/j.eswa.2016.04.032**