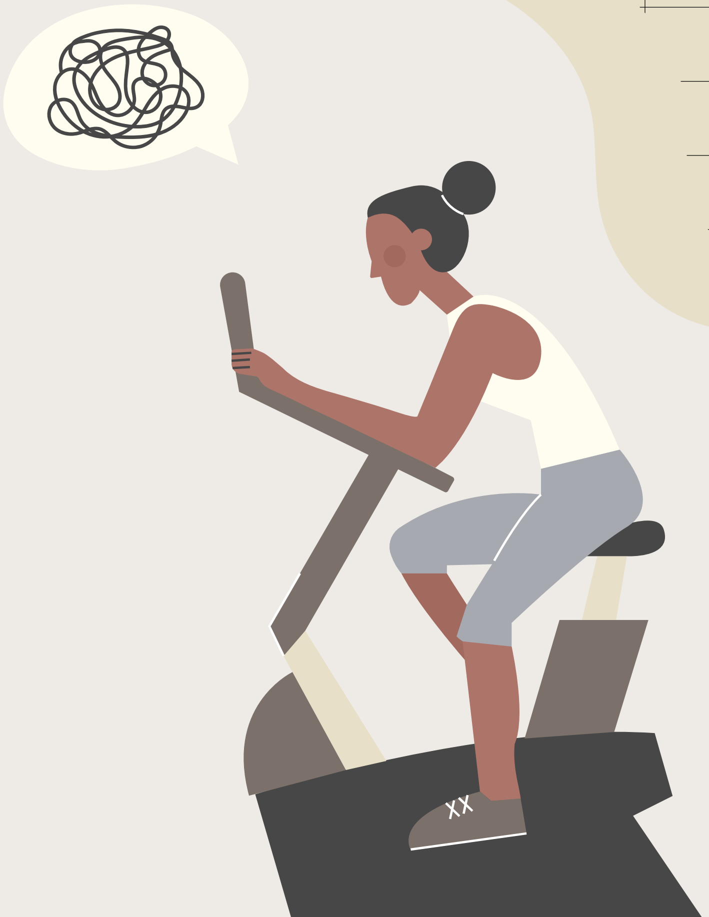
GenerZy: *“Small steps. Big energy. Keep your bar alive.”*

By: Lujain Osaily & Maryam Shanabli



“Every video game has an energy bar why shouldn’t your life have one?”

Are We Really Healthier, or Just Chasing Numbers??

Start

Endless numbers, little clarity.

Fitness and health apps flood us with **endless numbers of steps, calories, charts, scores**. Instead of clarity, people feel **confused and judged**

The myth

Society sells the myth of **"perfect health"** but chasing it often leaves people **burnt out, guilty, and unmotivated**.

The hidden cost of sitting

Long sitting hours quietly **drain our energy, weaken our health, and kill focus** yet trackers rarely show this hidden cost.

The emotional toll.

Constant tracking turns health into pressure, leaving people drained rather than encouraged.

End

“Redefining Health: Fun Over Perfection”

GenerZy turns daily activity into a game you want to play. Every step counts and earns points, badges, and nudges, making consistency rewarding. The live Energy Bar shows your energy fill and drain in real time, so you instantly know when to move, recharge, and how your efforts add up. Designed for a generation that values quick, visual, and motivating experiences, **GenerZy** replaces overwhelming numbers and charts with simple, empowering feedback. It’s not just a tracker t’s a fun, personal guide to staying active without stress or guilt.

Scope of Work – Data Preparation

PHASE ONE

Data Cleaning & Combination:
•Merged messy raw sensor files into **unified train/test CSVs**.
•Fixed missing values, duplicates, and inconsistent entries

Label Encoding:
•Converted activity names into **numeric IDs (1–6)** for ML.
•Ensured consistency across all datasets.

Train/Validation Split & Scaling:
•Split 80% train / 20% validation, **stratified by activity**.
•Applied **RobustScaler** to normalize features and handle outliers

Scope of Work – Feature Engineering & App-Class Mapping

PHASE TWO

•**Correlation & Spike Safety:** Keep features that matter, even if correlated, to capture rare sensor spikes.
•**Random Forest Importance:** Drop only low-impact features; top 95% contributors stay.
•**Final Output:** Reduced, clean feature matrices for train Val /test.

•**Purpose:** Map 6 activity types → 3 app-friendly classes:
•0 = SEDENTARY (SITTING + LAYING)
•1 = STANDING
•2 = ACTIVE (WALKING variants)
•**Validation:** Checked counts & class balance → mild imbalance handled via class weights.

Saved Artifacts:
•y_train_app.npy, Val, Val
•label_mapping_app.csv & app_class_names.csv

Impact / Why It Matters:
•No info lost: Rare events and key signals preserved.
•Simplified classes: Energy bar logic matches real-world movement.
•Reproducible & ready: All arrays, mappings, and checks saved for smooth modeling.

Scope of Work – Train Baseline Classifiers

Objective:
Train several baseline classifiers on the preprocessed, reduced feature set and evaluate their performance. This helps identify which model works best before hyperparameter tuning.

PHASE THREE

Data Preparation:
Load the **preprocessed, reduced features** and **3-class labels** (SEDENTARY, STANDING, ACTIVE).
Ensure shapes are consistent:
Feature arrays match the number of selected features.
Labels have the same number of rows as the corresponding feature arrays.

Check **class distribution** to confirm the mapping:
0 → SEDENTARY
1 → STANDING
2 → ACTIVE

Evaluation:
Predict on the **validation set**.
Metrics:
Accuracy → overall performance
Precision, Recall, F1-score → per class evaluation
Confusion Matrix → visual insight into misclassifications

PHASE THREE

Baseline Models:
We train several classifiers using default or reasonable parameters:

Model	Description
Random Forest	Robust to irrelevant features, handles non-linear relationships
Gradient Boosting	Sequential boosting, captures complex patterns
K-Nearest Neighbors (KNN)	Simple distance-based, good baseline for comparison
Support Vector Machine (SVM)	Effective for non-linear boundaries, can handle class imbalance

PHASE THREE

Model	Validation Accuracy	Notes / Highlights
Gradient Boosting	0.9905	Highest accuracy, stable predictions
Random Forest	0.9878	Very close to Gradient Boosting, all classes predicted well
KNN	0.9776	Slightly lower accuracy, may be slower on larger data
SVM	0.9742	Slightly lower accuracy, especially for STANDING

Validation Performance Summary:

Best Model: Gradient Boosting – slightly better accuracy and stable performance.
Runner-up: Random Forest – strong baseline, nearly identical performance.

Scope of Work – Gradient Boosting Hyperparameter Tuning

OBJECTIVE:
Improve the Gradient Boosting (GB) model by tuning key hyperparameters to achieve better generalization on validation and test sets.

PHASE FOUR

Data Preparation
Use the **final reduced features** and **3-class labels** from the previous phase.
Confirm that feature shapes match labels and check **class distributions**:
0 → SEDENTARY
1 → STANDING
2 → ACTIVE
Outcome: All datasets are aligned and ready for tuning.

Gradient Boosting Model
•Start with a **baseline GB classifier**:
- Sequential trees capture complex patterns.
- Parameters like number of trees, learning rate, and tree depth control model complexity.

Hyperparameter Grid
•Define a **search space** for tuning key hyperparameters:
- Number of trees, learning rate, tree depth
- Subsample fraction, minimum samples per split

•Keep the grid **reasonable** to balance tuning efficiency and model performance.

PHASE FOUR

Grid Search & Cross-Validation
Use **GridSearchCV** with **3-fold cross-validation** to find the optimal combination.
Each candidate model is trained multiple times to ensure reliable results.
Parallel processing accelerates the search.
Outcome: Best hyperparameters are automatically selected based on CV accuracy

Validation Evaluation
Evaluate the **best GB model** on the validation set.
Metrics include:
Accuracy → overall performance
Precision, Recall, F1-score → per-class evaluation
Confusion Matrix → visual insight into misclassifications
Example Result:
Validation Accuracy: 0.9946
All classes predicted accurately, showing excellent generalization.

Scope of Work – Test Set Evaluation

Goal:
Assess how well the tuned Gradient Boosting model generalizes to **completely unseen data**.

PHASE FIVE

Why Test Evaluation Matters
Validation accuracy can be **optimistic** because the model has seen this data indirectly during tuning.
The **test set** provides a **final unbiased estimate** of model performance.
Ensures that the model will perform well in real-world or deployment scenarios.

What We Measure
•**Accuracy** → overall correct predictions
•**Precision, Recall, F1-score** → per-class performance (SEDENTARY, STANDING, ACTIVE)
•**Confusion Matrix** → visual insight into which activities the model might confuse

Expected Outcome
•High accuracy close to the validation results indicates **good generalization**.
•For our tuned GB model, we expect:

- Accuracy > 0.99
- Minimal misclassifications between STANDING and ACTIVE classes
- Consistent performance across all three activity types

PHASE FIVE

How to Present Results
•Show a **confusion matrix heatmap** with actual vs predicted classes.
•Highlight **key metrics** in a table:

Metric	SEDENTARY	STANDING	ACTIVE
Precision	0.99	0.99	0.99
Recall	0.99	0.99	0.99
F1-Score	0.99	0.99	0.99
Support	2000	1800	2100

Key Takeaways
•The model maintains **high performance** on unseen data.
•Confirms the effectiveness of:

- **Feature selection** (top features capture most predictive power)
- **Hyperparameter tuning** (optimized Gradient Boosting model)

•Ready for deployment or further experimentation.

Scope of Work – Deployment

Goal:
Move the trained Gradient Boosting model from experimentation to **real-world usage**, either in **real-time** or **batch mode**.

PHASE SIX

In this phase, the **Generzy app** is deployed on **Hugging Face Spaces** using **Streamlit** for the Python front-end and **Docker** to containerize the application. This setup ensures that the **ML-powered energy bar** works seamlessly in a live environment, making it accessible from anywhere via a web link.

Team collaboration is managed through **community contributions**: teammates **fork the Space, edit their copy, and submit Pull Requests** for review. This approach allows the app to remain live while safely integrating updates, maintaining both functionality and version control during development.

Conclusion

Generzy app transforms everyday movement into a **fun, visual, and motivating experience**, encouraging healthy habits without overwhelming metrics. By deploying it on **Hugging Face Spaces** with **Streamlit and Docker**, the app is fully accessible online, integrates ML-powered energy tracking, and supports smooth team collaboration via Pull Requests. This approach demonstrates how **gamification, modern tools, and innovative deployment** can make health tracking engaging, scalable, and collaborative empowering users to turn small daily actions into meaningful energy and progress.

THANK YOU