DSA-210 FINAL REPORT

**Başar Erses**

**31904**

Battery Efficiency Project

Sabancı University

2024-2025 Spring Term

Supervised by: Özgür Asar

# Content

This report explores battery efficiency across various devices and tasks by analyzing system-level metrics such as brightness, CPU utilization, temperature, memory usage, and battery health. Using machine learning models, statistical tests, clustering, and trend analysis, this project aims to uncover patterns and deliver actionable insights for optimizing energy consumption and improving device longevity.

# INTRODUCTION

Modern computing devices operate under diverse workloads with power consumption directly tied to system usage patterns. This project aims to model and analyze how different usage contexts affect battery life using a multi-faceted data science approach. Data collected over multiple weeks across phones, tablets, and laptops captures real-world usage under tasks such as coding, streaming, gaming, and idle periods.

# WHAT DID I DO?

The dataset includes timestamped logs of device status including screen brightness, CPU and memory usage, task type, temperature, and battery statistics. Data was cleaned and transformed, followed by visual and statistical analysis to reveal patterns. Regression models predicted battery life based on feature values, while classification and clustering techniques helped categorize usage behaviors.

Hypothesis testing was applied to validate the influence of task type on battery drain, and SHAP explainability techniques were used to interpret model results. Clustering algorithms such as KMeans uncovered user profiles based on consumption. Finally, time-based trends tracked degradation in battery health.

# KEY OBSERVATIONS AND ANALYSIS

Brightness, CPU usage, and temperature were strongly negatively correlated with battery life.

**Feature Correlation:** Brightness and CPU utilization consistently showed strong negative correlations with battery life. Higher brightness levels and intensive CPU usage were associated with rapid battery drain. Device temperature also had a mild negative effect, especially during prolonged usage sessions.

**Regression Modeling:** Linear Regression provided a reasonable baseline, but Ridge Regression improved generalization by penalizing overfitting. Gradient Boosting delivered the best performance, capturing complex, non-linear patterns and interactions between system metrics.

**Classification Performance:** The Random Forest Classifier achieved high accuracy in predicting task types (e.g., gaming, coding, streaming) based on usage statistics. Label encoding and stratified train-test splitting ensured balanced performance across classes.

**Model Explainability (SHAP):** SHAP values revealed the inner logic of the Random Forest model. CPU usage, memory usage, and temperature were identified as the most influential features, providing intuitive insight into how system load affects energy consumption.

**Clustering Insights:** KMeans clustering grouped the data into three distinct behavioral profiles: light users (idle/coding), moderate users (streaming/office work), and heavy users (gaming/video editing). These clusters were confirmed by distribution plots showing the variation in battery life and system usage across groups.

**Trend and Longitudinal Analysis:** Battery health showed a consistent downward trend over time, especially for devices used for resource-intensive tasks. The degradation rate was steeper for gaming and video editing compared to idle or light usage periods.

**Statistical Hypothesis Testing (ANOVA):** One-way ANOVA validated that task type significantly influenced average battery life. The p-value was well below 0.05, rejecting the null hypothesis and confirming that usage context has a statistically measurable impact on power consumption.

- SHAP analysis confirmed CPU utilization and brightness as the most influential predictors.

- Regression models (Linear, Ridge, Gradient Boosting) offered accurate predictions with cross-validation.

- Classification using Random Forest achieved reliable task categorization.

- KMeans clustering revealed behavioral segments (light, moderate, and heavy usage profiles).

- ANOVA confirmed statistically significant impact of task type on battery life.

- Battery health trends varied by task intensity and device type over time.

# UNDERSTANDING THE VARIABILITY

While individual metrics had measurable impacts, real-world battery performance depended on the interaction of multiple features. For example, high CPU usage combined with maximum brightness had a compounded negative effect. On the other hand, idle states with moderate memory usage and low temperature preserved battery life.

# How could it be improved?

Larger datasets from more device brands, and continuous logging with advanced monitoring tools could improve the precision and scalability of this project. Time constraints limited deeper exploration into multivariate interaction effects and the integration of adaptive energy-saving algorithms.

# Conclusion

This project highlights the importance of optimizing energy consumption based on task, hardware condition, and system load. It demonstrates how data science can guide both users and designers in creating power-efficient environments.

In the future integrating real-time prediction models to enable smart battery management or develop adaptive profiles tailored to user behavior. With better monitoring and ongoing adjustment, energy efficiency can be significantly improved across a wide range of devices.