|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| DSA-210 FINAL REPORT | | | | |
|  |  | | |  |
|  | | FURKAN YİĞİTBAKIM |  | |
|  | | 32631 EFFECT OF NUMBER OF LECTURE HOURS ON  PICK-UP NUMBER  **SUPERVISED BY** **ÖZGÜR ASAR** |  | |

**SABANCI UNIVERSITY**

|  |
| --- |
|  |

**WHAT'S IN THIS REPORT?**

This report summarizes the entire workflow of my term project: data collection and cleaning, exploratory data analysis, hypothesis testing, machine learning models and results, and key takeaways.

**Parameters in the Report:**

* **LectureHours (hours)**: Total daily academic workload
* **TotalPickups (count)**: Total number of phone pick-ups
* **SocialMediaPickups (count)**: Pick-ups for social media apps
* **CommunicationPickups (count)**: Pick-ups for communication apps
* **TotalScreenTime (minutes)**: Total screen time recorded
* **SleepDuration (minutes)**: Total sleep duration for each day
* **CaffeineIntake (mg)**: Daily caffeine consumption
* **IsHighCaffeineDay (binary)**: 1 if caffeine intake > 300 mg, 0 otherwise
* **Gym (binary)**: 1 if attended gym that day, 0 if not
* **IsWeekend (binary)**: 1 for weekends, 0 for weekdays
* **Temperature (°C)**: Daily average temperature
* **Condition (categorical)**: Weather condition (e.g., Sunny, Rainy)
* **IsExamDay (binary)**: 1 for exam days, 0 for non-exam days

**INTRODUCTION**

The project explores how my daily academic workload (specifically lecture hours) influences my phone usage patterns — focusing on total phone pickups, social media usage, and communication app usage. I manually recorded data for three months, tracking contextual factors like sleep, caffeine intake, gym activity, and weather. I also created a custom binary feature called IsHighCaffeineDay to capture days when caffeine intake exceeded 300mg, which I suspected might also affect phone habits.

**WHAT DID I DO?**

**Data Collection & Preparation**

* I logged my daily behavior and habits in an Excel file (data.xlsx).
* The dataset included: lecture hours, total phone pickups, social media and communication pickups, sleep duration, caffeine intake, gym attendance, and weather information.
* I cleaned the data, removed missing values, created the IsHighCaffeineDay and IsWeekend binary features, and renamed columns for easier analysis.

**Exploratory Analysis & Visualizations**

* I explored correlations between variables using heatmaps and scatter plots.

metin, ekran görüntüsü, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

* LectureHours showed a positive correlation with both social media pickups and communication pickups, suggesting a link between academic load and phone usage. I also examined how behavior changed over time with line plots.

metin, çizgi, ekran görüntüsü, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, çizgi, ekran görüntüsü, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Key Finding:**

The best performance was achieved with just two features: LectureHours and IsHighCaffeineDay. Using only these, the Random Forest model reached an R² of 0.361 for predicting total pickups.

I also predicted social media and communication pickups using the same two features:

| **Target** | **R² Score** |
| --- | --- |
| Social Media Pickups | 0.689 |
| Communication Pickups | 0.464 |
|  |  |

This confirmed that these two simple features captured a significant part of my phone behavior.

**Model Comparison:**

To show why some models weren’t used, I compared all models using only the two best features:

* Linear Regression: R² = 0.144 | MAE = 29.58 | RMSE = 46.35
* Random Forest: R² = 0.361 | MAE = 28.06 | RMSE = 40.05
* K-Nearest Neighbors: R² = 0.266 | MAE = 26.23 | RMSE = 42.92
* Support Vector Regressor: R² = -0.128 | MAE = 37.07 | RMSE = 53.21

Random Forest had the best balance of R² and errors. SVR had a negative R², failing to model the data at all. Decision Tree had similar results to Random Forest, but Random Forest was more robust.

**Conclusion:**

This project helped me see how academic load and caffeine consumption shape my phone usage habits.

Key takeaways:

* Lecture hours and caffeine are the biggest predictors of phone usage.
* Minimal features can still capture meaningful behavior if chosen carefully.
* Even in small, personal datasets, the right model and feature engineering can uncover real insights.

Beyond the data, this project made me more aware of my digital habits and how they shift with academic pressure. It also taught me how to combine manual tracking, exploratory data analysis, and machine learning to turn simple logs into powerful observations. Overall, it was a journey not just to learn data science tools, but also to better understand myself.

metin, ekran görüntüsü, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu

**CAN IT BE BETTER?**

Of course!

* The dataset was small and personal, so results might not generalize to everyone.
* In future work, I could track stress levels, motivation, or other qualitative factors.
* Collaborating with more students would create a larger, more robust dataset.

**FINAL WORDS**

This wasn’t just about numbers, it was about understanding my own patterns and using data to make smarter choices. It was also a chance to learn practical data science skills: from data cleaning and visualization to model evaluation.

More than that, this project pushed me to look at my daily habits in a new light. It showed me how even small pieces of data — like lecture hours or caffeine intake — can tell a powerful story about our daily lives and how we adapt to academic pressures.

By combining data analysis with personal curiosity, I learned that data science is not just about predicting numbers, it’s about making sense of them and connecting them to real, human experiences.

This project has made me more mindful of how I balance my studies, lifestyle, and technology use. It’s also inspired me to keep exploring how data can help us make more informed decisions, both in school and beyond.

Thanks for reading!