Discovering Short-Term Fatigue-Management Patterns: Power Naps vs. Coffee Across Sleep, Duration, and Side Effects

Kofi Ofori-Acquah
CS 4412: Data Mining
Kennesaw State University
Email: koforiac@students.kennesaw.edu

***Abstract*—This project proposes a data mining project focused on *pattern discovery* rather than prediction accuracy, using the Kaggle dataset *Power Nap vs Coffee Effectiveness*. The dataset represents simulated short-term response data for two fatigue-management interventions: coffee and power naps, including prior sleep duration, intervention duration, alertness scores before and after, productivity and mood ratings, and side effects. The project's main goal is to discover patterns within the dataset: (1) segments of individuals responding differently to each intervention, (2) association rules between sleep/intervention conditions and positive outcomes and side effects, and (3) unusual response patterns. The project's plan includes a variety of mining techniques: clustering, association rule mining, anomaly detection, and PCA for visualization, which will provide human-readable results on patterns within the dataset's outcomes.**

***Index Terms*—data mining, clustering, association rules, anomaly detection, PCA, exploratory data analysis**

I. Introduction

Selecting between a power nap and coffee is an everyday decision that individuals face for managing fatigue. Although predictive modeling can forecast outcomes, this project's focus is on **discovery**, where we want to discover patterns and relationships that explain why outcomes vary across sleep levels, intervention settings, and side effects.

II. Dataset Description

A. Source and Size

Dataset: *Power Nap vs Coffee Effectiveness Dataset* (Kaggle).
URL: https://www.kaggle.com/datasets/prince7489/power-nap-vs-coffee-effectiveness-dataset
Format: single CSV file.
Size: 500 rows $\times$ 11 columns ($\approx$ 6.4KB).

B. What the Data Represents

This dataset represents simulated fatigue management sessions, one per row, where each row represents one participant. There are demographic factors, prior sleep, type and duration of interventions, and alertness levels before and after interventions, as well as self-reported productivity/mood outcomes. There are also side effects reported, which can be missing.

C. Attributes

Columns in the data are:

- `participant_id` (int), `age` (int), `occupation` (categorical)
- `sleep_hours_previous_night` (float)
- `intervention_type` (categorical: Coffee, Power Nap)
- `intervention_duration_minutes` (int)
- `alertness_score_before` (int), `alertness_score_after` (int)
- `productivity_rating` (int, 1–10), `mood_rating` (int, 1–10)
- `side_effects` (categorical; can be missing)

D. Planned Derived Features and Data Quality Notes

Derived feature: **alertness_delta** = `after` − `before`. Other discretized features will be created for association rules. Data quality considerations: `side_effects` can be missing, and there are considerations for categorical normalization, and because it is simulated, it is worth noting that patterns are reported as discovered within the data, rather than being descriptive of the real world.

III. Discovery Questions

A. DQ2: Conditional Effectiveness Patterns

What are the sleep level, intervention choice/duration, and occupation combinations strongly associated with high alertness improvement and/or high productivity/mood?
Motivation: extracting human-readable conditions using association rules (e.g., conditions $\Rightarrow$ high delta/high productivity) and ranking by support/confidence/lift.

B. DQ3: Side Effects and Anomalous Profiles

What conditions are associated with side effects, and are there anomalous sessions (unexpectedly low/negative or unusually high improvements) compared to similar profiles?
Motivation: discovering trade-offs and identifying unusual cases for deeper explanation.

TABLE I
SCHEMA OVERVIEW OF THE POWER NAP VS. COFFEE EFFECTIVENESS DATASET

Attribute	Type	Description
participant_id	Integer	Unique participant identifier
age	Integer	Participant age
occupation	Categorical	Participant occupation
sleep_hours_previous_night	Numeric	Prior sleep duration
intervention_type	Categorical	Coffee or Power Nap
intervention_duration_minutes	Integer	Intervention duration
alertness_score_before	Integer	Alertness before intervention
alertness_score_after	Integer	Alertness after intervention
productivity_rating	Integer	Self-reported productivity (1–10)
mood_rating	Integer	Self-reported mood (1–10)
side_effects	Categorical	Reported side effects

IV. PLANNED TECHNIQUES AND ANALYSIS APPROACH

The project will employ various techniques from different categories of data mining: clustering, association rules, anomalies, and dimensionality reduction. The techniques are followed by preprocessing and exploratory data analysis.

A. Preprocessing and EDA

- Data profiling
- Data cleaning: handling missing values in column `side_effects` (introduce “None/Not Reported” category)
- Data encoding: one-hot encoding for categorical variables
- Data scaling: standardization for distance-based algorithms
- Data engineering: create new feature `alertness_delta` and bucketed variables
- Exploratory data analysis: group comparisons based on intervention type and duration; correlations for numeric variables

B. Clustering (for DQ1)

- K-Means clustering on a set of numeric features (sleep level, duration, before/after, delta, productivity, mood)
- Hierarchical clustering as a secondary approach for better interpretability
- Cluster evaluation based on interpretability and stability
- Descriptive profiles for clusters

C. Association Rule Mining (for DQ2 and DQ3)

Association rules mining will be performed using Apriori or FP-Growth algorithm after discretization of numeric features into ranges (sleep bucket, duration bucket, delta tier, productivity tier, mood tier) and inclusion of categorical features (occupation, intervention type, side effects). Association rules will be evaluated and ranked using support, confidence, and lift metrics.

D. Dimensionality Reduction (PCA) (supports DQ1, DQ3)

Dimensionality reduction using PCA algorithm will be performed on the standardized numeric feature space to enable visualization of cluster separations and anomalies in 2D.

E. Anomaly Detection (for DQ3)

Anomalies will be detected using statistical thresholds for individual features (`alertness_delta`, productivity, mood) and optionally local approaches for detecting points that are far from nearby points.

V. PLANNED ANALYSIS PIPELINE

Data Load $\rightarrow$ Cleaning/EDA $\rightarrow$ Feature Engineering (delta + bins) $\rightarrow$ $\{$Clustering + PCA$\}$ $\rightarrow$ $\{$Association Rules (overall + within clusters)$\}$ $\rightarrow$ $\{$Anomaly Detection$\}$ $\rightarrow$ Interpretation + Visuals

VI. PRELIMINARY TIMELINE

M2: Data load, cleaning, EDA, feature engineering; initial clustering; baseline association rules.
M3: Refine clustering (compare clustering methods and parameters); refine association rules thresholds; PCA visualization; anomaly detection and explanations.
M4: Synthesize strongest discovered patterns; finalize visuals and tables; reproducibility and repository polish; final report writing.

VII. ANTICIPATED CHALLENGES

- Discretization thresholds for meaningful association rules
- Quantitative validation versus interpretability
- Discovery framing for simulated dataset without causal claims

VIII. GITHUB REPOSITORY PLAN

Repository: https://github.com/kof-coder/cs4412-project.git
Planned repository structure:

- `README.md`: title, discovery-oriented summary, dataset source link, author/team
- `docs/`: proposal PDF
- `data/`: dataset CSV or instructions/.gitignore if needed

IX. Conclusion

This proposal outlines the development of a data mining project that revolves around the discovery of the relationship between naps and coffee in the context of short-term alertness, productivity, mood, and side effects. The data mining process, which entails clustering, association rules, anomaly detection, and PCA visualization, is intended to result in the generation of patterns and responses that are relevant to the dataset.

References

[1] M. J. Zaki and W. Meira Jr., *Data Mining and Machine Learning: Fundamental Concepts and Algorithms*. Cambridge University Press, 2014.

[2] P. 7489, "Power Nap vs Coffee Effectiveness Dataset," Kaggle, accessed 2026. [Online]. Available: https://www.kaggle.com/datasets/prince7489/power-nap-vs-coffee-effectiveness-dataset