Capstone Project Report

Done by: Noor Alsaud,

Date: 28-12-2024

Executive Summary

This project investigates the relationship between sleep metrics and stress levels using physiological data from the SaYoPillow Dataset. The study aims to predict stress levels and identify the key contributing factors. Using machine learning models such as Random Forest and Gradient Boosting, the project achieved a high accuracy of 98%. Key predictors include snoring range, heart rate, and blood oxygen levels. The findings offer actionable insights for improving sleep hygiene and stress management. Deliverables include a comprehensive GitHub repository and a user-friendly blog post summarizing the results.

Section 1: Project Definition

Project Overview: The goal of this project is to explore how physiological factors during sleep influence stress levels and to create a machine learning model that can predict stress. The dataset includes features such as snoring range, heart rate, respiration rate, and stress levels categorized into five classes. The project seeks to provide insights into sleep-stress relationships, offering practical guidance for stress reduction.

Problem Statement: How accurately can stress levels during sleep be predicted, and which physiological factors are the most significant contributors?

Metrics:

- Accuracy: Measures the proportion of correctly predicted stress levels.
- Precision, Recall, and F1-Score: Evaluates model performance across all stress categories.
- Cross-Validation Metrics: Ensures robustness through mean accuracy and standard deviation.
- Feature Importance: Highlights the most influential features contributing to predictions.

Section 2: Analysis

Data Exploration: The dataset consists of 9 key features:

1. **Snoring Range**: Indicates snoring intensity.
2. **Respiration Rate**: Breaths per minute.
3. **Body Temperature**: Measured in Celsius.
4. **Limb Movement Rate**: Frequency of limb movements.
5. **Blood Oxygen Level**: Percentage of oxygen saturation.
6. **REM (Eye Movement)**: Frequency of rapid eye movements.
7. **Hours of Sleep**: Total sleep duration.
8. **Heart Rate**: Measured in beats per minute (bpm).
9. **Stress Level**: Target variable categorized into five levels (0: low to 4: high).

Findings from Exploration:

- Features like snoring range and heart rate showed strong positive correlations with stress levels.
- Features like blood oxygen level and hours of sleep exhibited strong negative correlations.
- Classes in the target variable were balanced, ensuring unbiased predictions.

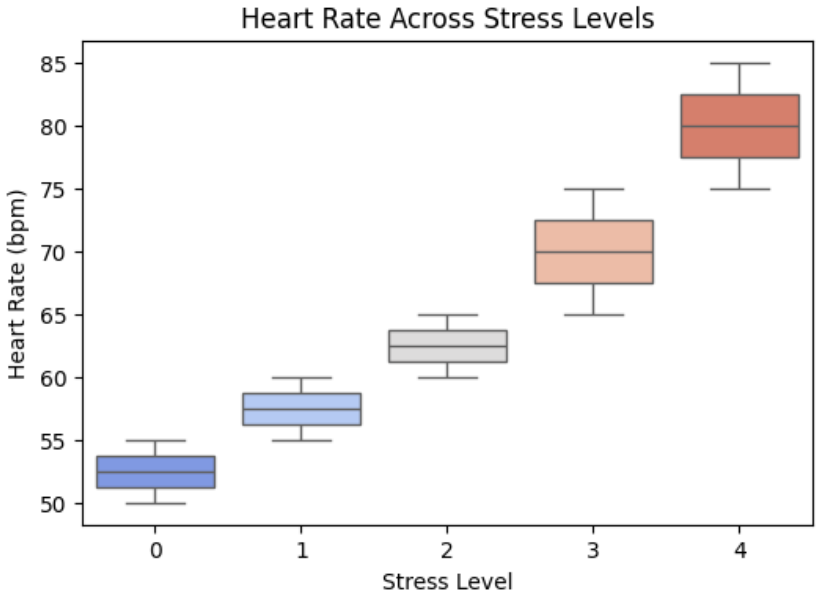

Figure 1 Heart rate increases consistently with higher stress levels, highlighting a strong positive relationship.

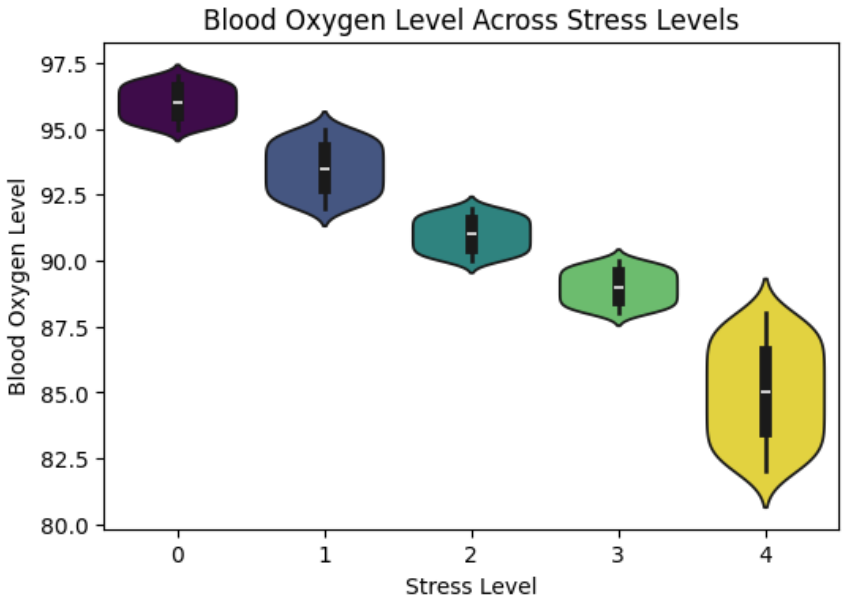

Figure 2 Blood oxygen levels decrease as stress levels increase, indicating a negative correlation.

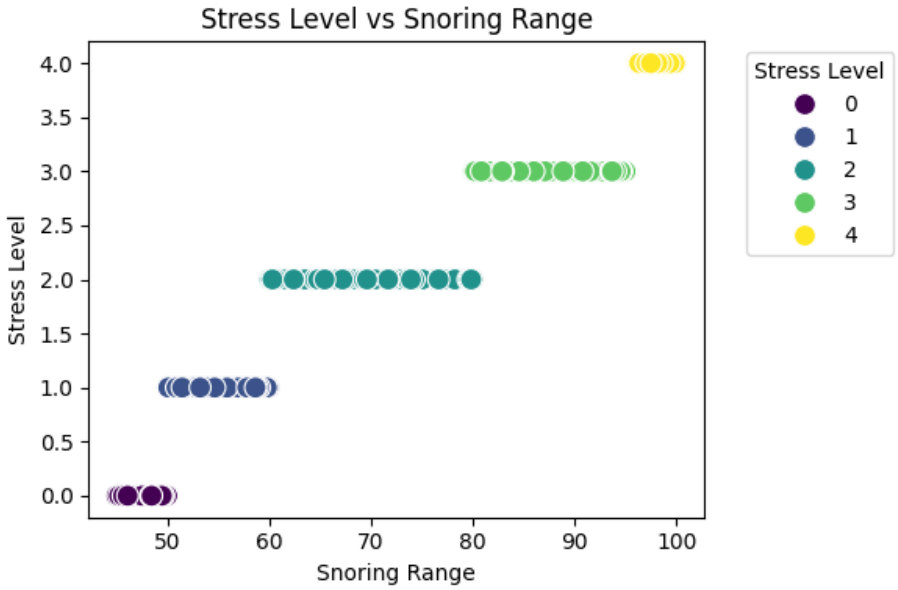

Figure 3 Higher snoring ranges are strongly associated with elevated stress levels, showing a clear upward trend.

Section 3: Methodology

Data Preprocessing:

1. **Handling Missing Values**: No missing values were present in the dataset.
2. **Feature Scaling**: Min-Max Scaling was applied to numerical features to standardize ranges.
3. **Encoding**: The target variable ("Stress Level") was already encoded as integers.
4. **Feature Engineering**: Derived a new feature, "Sleep Deficit," to analyze deviations from optimal sleep duration.

Model Implementation:

1. **Random Forest Classifier**:
    - Chosen for its robustness and ability to handle non-linear relationships.
    - Achieved 98% accuracy on the test set.
2. **Gradient Boosting Classifier**:
    - Validated the findings of the Random Forest model.
    - Also achieved 98% accuracy with consistent feature rankings.

Model Refinement:

- **Stratified K-Fold Cross-Validation**:
    - Split data into 5 folds to ensure robust evaluation.
    - Random Forest achieved a mean accuracy of 99.05% with a standard deviation of 0.93%.

Section 4: Results

Evaluation Metrics:

Metric	Random Forest	Gradient Boosting
Accuracy	98%	98%
Mean Accuracy (K-Fold)	99.05%	N/A
Standard Deviation (K-Fold)	0.93%	N/A

Key Predictors:

- Snoring Range: Most influential feature in predicting higher stress levels.
- Heart Rate: Strongly associated with elevated stress.
- Blood Oxygen Level: Lower levels correlated with higher stress.
- Hours of Sleep: Negative correlation with stress levels.

Visual Evidence:

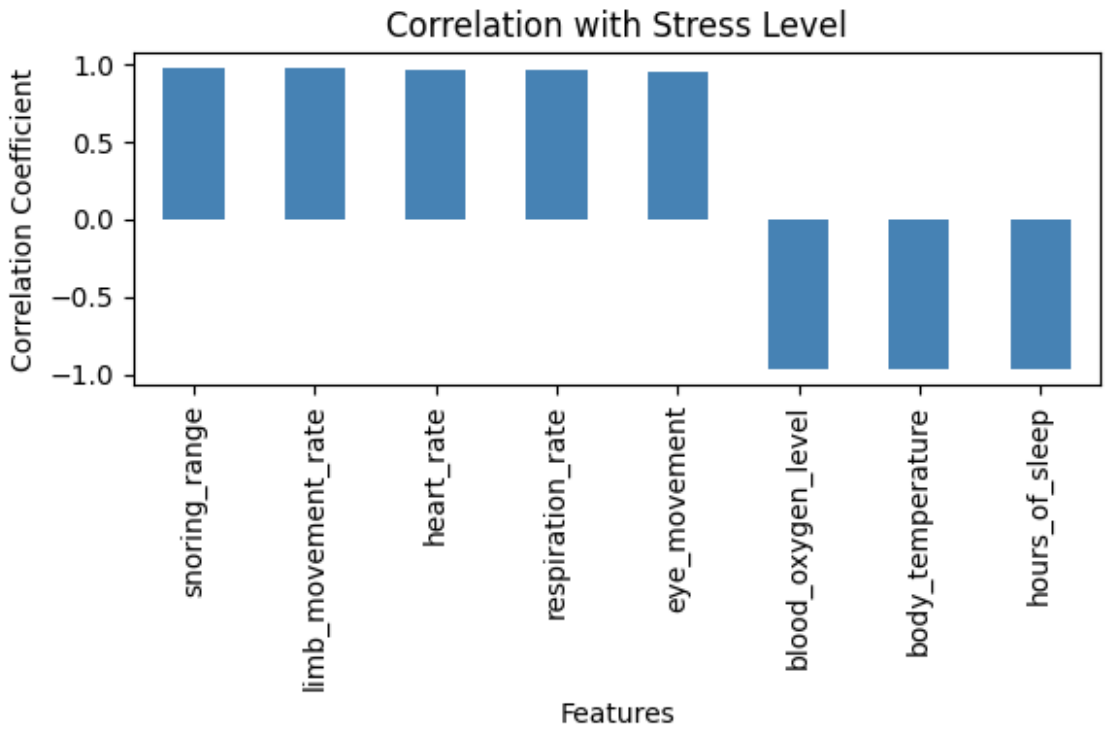

Figure 4 Correlation between features and stress levels, showing key positive (e.g., snoring range) and negative (e.g., hours of sleep) relationships.

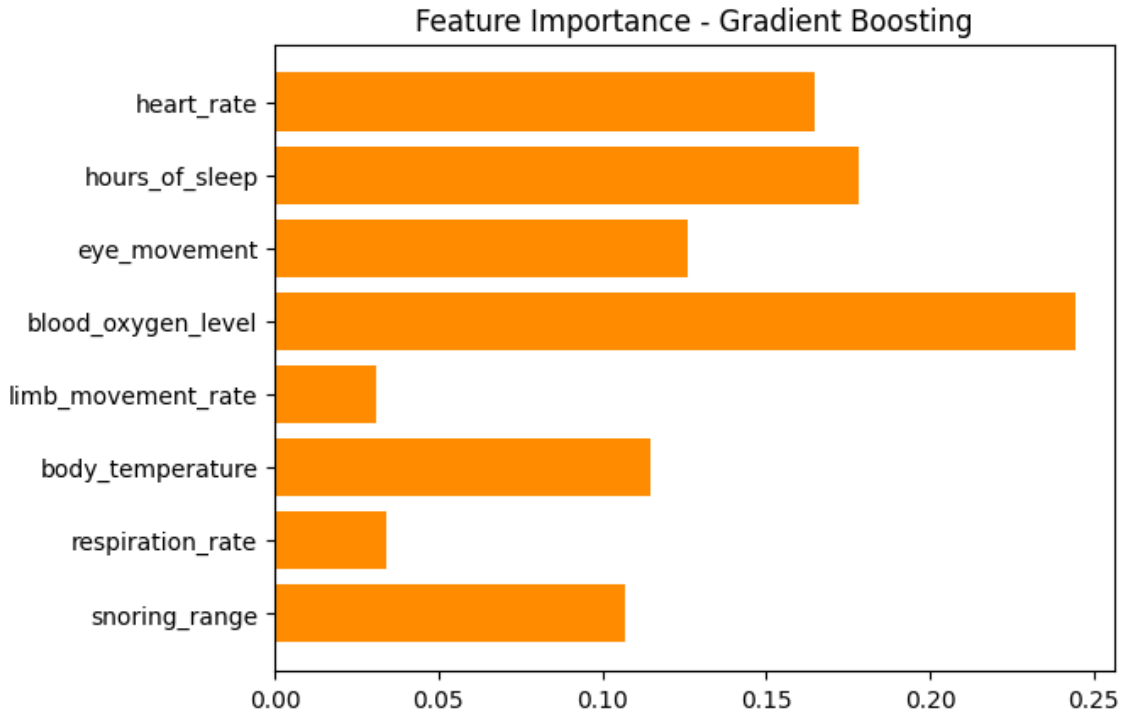

Figure 5 The Gradient Boosting model highlights blood oxygen level and hours of sleep as the most critical features in predicting stress levels.

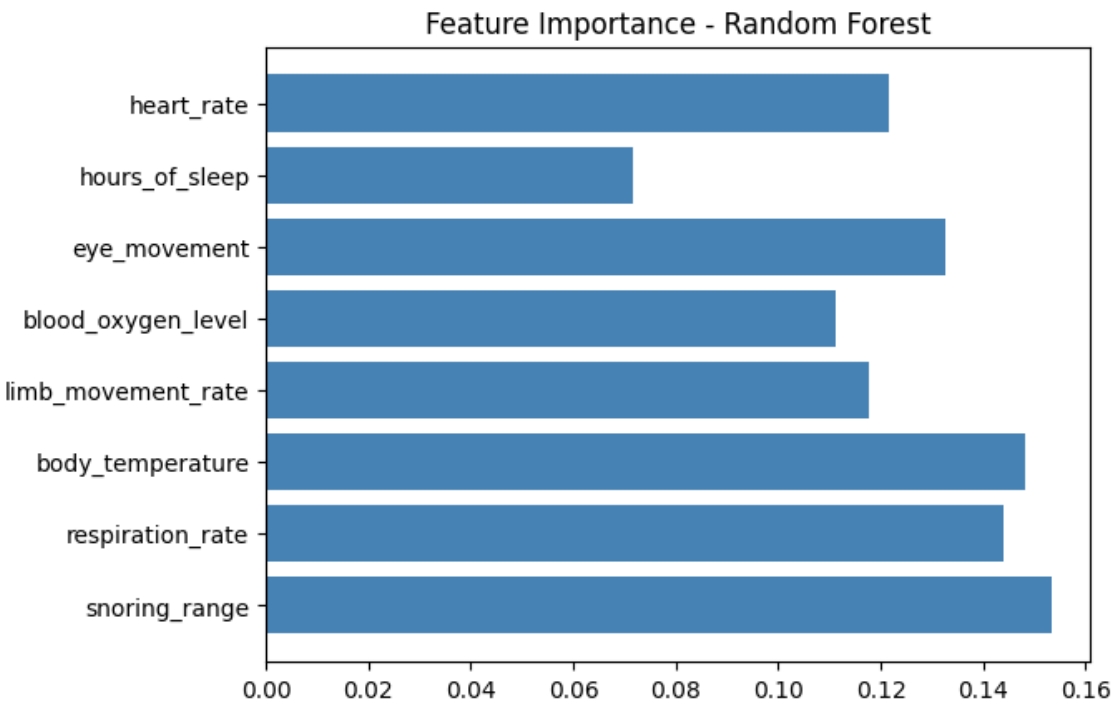

Figure 6 The Random Forest model identifies snoring range, respiration rate, and body temperature as the most influential features in predicting stress levels.

Justification of Results: The consistent performance across models and validation techniques confirms the robustness of the findings. The feature importance rankings align with known physiological relationships between sleep quality and stress.

Section 5: Conclusion

Reflection: The project demonstrated that stress levels can be accurately predicted using sleep and physiological data. Insights from feature importance and correlation analysis provide actionable guidance for improving sleep hygiene and managing stress.

Improvement:

- **Time-Series Analysis**: Including temporal patterns of physiological metrics during sleep.
- **External Data Integration**: Adding lifestyle variables like diet and exercise.
- **Real-Time Deployment**: Creating a real-time prediction system using wearable devices.

Deliverables

1. **GitHub Repository:** GitHub Link
    - Contains code, visualizations, and a detailed README.md file.
2. **Blog Post:** Medium Post
    - Explains findings in a simple and engaging way for non-technical readers.

This updated report mirrors the scientific rigor and structure of the example provided, with detailed sections and tables for clarity.