1.TITLE

Sleep Quality Prediction Using Machine Learning

Name: *Bangaramma Kasani*

Department: *Artificial Intelligence & Data Science(AI&DS)*

College/University: Vishnu Institute of Technology

Project Guide: *(Guide Name)*

Year: *(2025-2026)*

2. Abstract

This project focuses on predicting sleep quality using machine learning. The dataset contains information about sleep patterns and lifestyle factors such as sleep hours, screen time, caffeine intake, exercise, stress level, gender, etc. After performing data cleaning and preprocessing, multiple machine learning models were trained and evaluated. Random Forest Classifier provided the best accuracy of 99%. The trained model is used to predict sleep quality for new data input. This project helps in understanding sleep patterns and improving health awareness.

3. Introduction

Sleep is one of the most important factors for maintaining physical and mental health. Poor sleep quality can cause stress, low productivity, mood changes, and health problems. Predicting sleep quality using machine learning helps people understand the reasons behind poor sleep and take preventive steps. This project uses machine learning techniques to analyze data and predict whether a person's sleep quality is good or poor.

4. Problem Statement

The main goal of this project is to build a machine learning model that can predict sleep quality based on lifestyle and sleep-related factors.

5. Objectives

- To collect and analyze sleep-related data

- To perform data cleaning and preprocessing

- To train machine learning models

- To evaluate the performance of different models

- To select the best model for prediction

- To predict sleep quality for new input data

6. Dataset Description

The dataset contains 700 records of sleep and lifestyle information of individuals. It includes features such as sleep hours, screen time, caffeine intake, exercise, stress level, gender, and sleep quality. The data is stored in Excel format. The dataset is used for training and testing the machine learning model.

7. Data Cleaning and Preprocessing

Data cleaning includes removing unwanted columns, checking for missing values, and correcting incorrect data. Preprocessing includes converting categorical values into numeric format and scaling the numerical features. Only the required columns used for prediction are converted into numeric values. Scaling is performed using StandardScaler to improve model performance.

8. Machine Learning Models Used

Different models were trained to find the best one:

- Logistic Regression

- Decision Tree Classifier

- Random Forest Classifier

- K-Nearest Neighbors (KNN)

The Random Forest model provided the best accuracy and was selected as the final model.

9. Model Evaluation

The models were evaluated using accuracy score. The results are:

Model	Accuracy
Logistic Regression	94%
Decision Tree	100%
Random Forest	99%
KNN	86%

Random Forest was chosen because it gives high accuracy and works well for prediction.

10. Prediction

The trained Random Forest model is used to predict sleep quality for new input data. Based on the input features, the model outputs whether the sleep quality is good, average, or poor. This prediction helps users understand their sleep pattern and improve their sleep quality.

11. Tools and Technologies Used

- Python

- Pandas

- NumPy

- Scikit-learn

- VS Code

- GitHub

12. Conclusion

This project successfully predicts sleep quality using machine learning techniques. The Random Forest model achieved high accuracy and can predict sleep quality effectively. The project demonstrates the application of machine learning in health-related problems and helps people understand sleep patterns.

13. Future Scope

- Collect data from wearable devices for real-time prediction

- Increase dataset size for better accuracy

- Deploy the model as a web or mobile application

- Add more features like diet, medical history, and lifestyle

14. References

- Scikit-learn Documentation

- Pandas Documentation

- Research articles on sleep quality

- Kaggle dataset sources