README

# Author

YE Hongbo, hongbo.ye@edu.dsti.institute

YANG Yingpeng, yingpeng.yang@edu.dsti.institute

BALAKRISHNAN Abishekk, abishekk.balakrishnan@edu.dsti.institute

# Introduction

This project aims to explore the sleep patterns of various animal species and build predictive models for two key sleep metrics: TotalSleep and Dreaming. The dataset used in this project contains information about different animal species including their body weight, brain weight, lifespan, gestation period, and various categorical features related to their behavior, habitat, and conservation status.

# Files Included

1. Python ML Final Project.ipynb: This Jupyter Notebook contains the entire code for the project including data preprocessing, exploratory data analysis, feature engineering, and predictive modeling.
2. README.md: This README file provides an overview of the project, its objectives, and the files included.
3. sleep\_merged.tsv: This tab-separated values (tsv) file contains the raw data used for analysis.
4. requirements.txt: This file lists all the Python libraries and dependencies required to run the project code.

# Installation and Setup

1. Ensure you have Python installed on your system.
2. Clone or download the project repository from GitHub link.
3. Open the Jupyter Notebook Python ML Final Project.ipynb in Jupyter Notebook or JupyterLab.

# Instructions for Running the Code:

1. Open the Jupyter Notebook Python ML Final Project.ipynb.
2. Execute each code cell sequentially by either clicking on "Run" or using the keyboard shortcut Shift+Enter.
3. Follow the comments in the code cells for understanding the purpose and functionality of each code block.
4. Make sure to update file paths if you're using a different directory structure or file names.

# Project Structure:

1. Package Installation & Data Upload: This section installs necessary Python packages and imports the dataset for analysis.
2. Data Checking: Performs initial checks on the dataset including checking for missing values, duplicates, and data types.
3. Exploratory Data Analysis (EDA): Conducts exploratory analysis to understand the distribution and relationships between variables. Includes visualizations such as histograms, boxplots, scatter plots, and correlation matrices.
4. Feature Engineering: Preprocesses the data by handling missing values, encoding categorical variables, and selecting relevant features using mutual information scores.
5. Modeling for TotalSleep and Dreaming: Builds predictive models using Random Forest, Gradient Boosting, and Support Vector Regression algorithms. Evaluates model performance and selects the best model based on Root Mean Squared Error (RMSE). Visualizes feature importance and SHAP (SHapley Additive exPlanations) values for model interpretation.

# Conclusion:

This project demonstrates the process of exploratory data analysis and predictive modeling for understanding and predicting sleep patterns in animals. By analyzing various features and building machine learning models, insights can be gained into factors influencing total sleep duration and dreaming behavior across different animal species.