

Sleep Disorder Classification

— From-Scratch Model Comparison

This repository trains and compares four classifiers implemented **from scratch** (no `sklearn` models used):

- K-Nearest Neighbors (KNN)
- Decision Tree
- Random Forest (ensemble of Decision Trees)
- Linear SVM (one-vs-rest, primal hinge loss with SGD)
- Logistic Regression

The pipeline:

- Normalizes column names to `snake_case`
- Maps target labels (`None`, `Insomnia`, `Sleep Apnea`) → numeric codes (0,1,2)
- One-hot encodes categorical features
- Standardizes numeric features (fit on training set)
- Performs a stratified train/test split
- Trains all models on the same split
- Evaluates models and produces comparative plots and a small HTML report

Files

- `models/knn.py` — KNN implementation
- `models/decision_tree.py` — Decision tree implementation
- `models/random_forest.py` — Random forest + permutation importance
- `models/svm.py` — linear SVM (from-scratch) OvR
- `models/logistic_regression.py` - Logistic Regression(from scratch)
- `utils.py` — preprocessing, custom classification report, scaler
- `run.py` — main script to train all models, plot, and save outputs/ + artifacts/
- `requirements.txt` — Python dependencies

Quick start

1. Clone repo:

```
git clone https://github.com/555vedant/ML_Project.git
cd ML_Project
```

2. Create virtualenv and install:

```
python -m venv venv
source venv/bin/activate    # Windows: venv\Scripts\activate
pip install -r requirements.txt
```

3. Run:

```
python run.py --data Sleep_health_and_lifestyle_dataset.csv
```