**Project Report**

**Sleep Disorder Prediction and Well-being Enhancement**

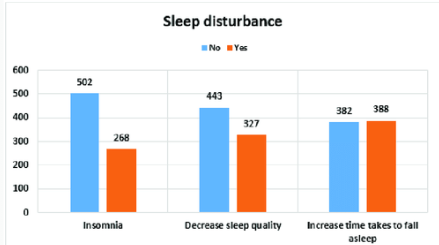
**Kandakatla Vaishnavi**

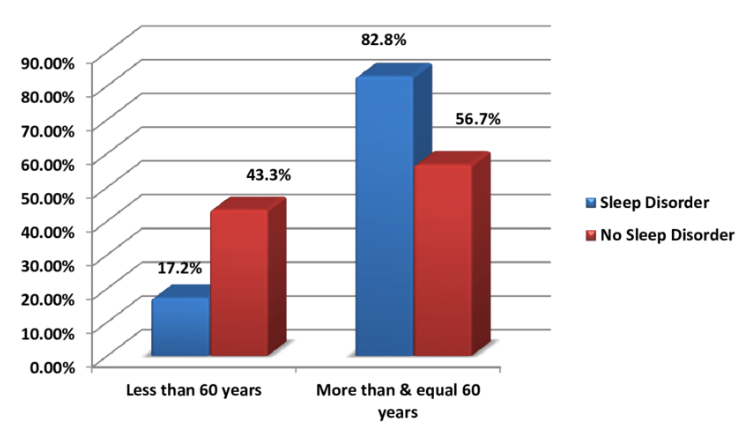
**U00918787- Group 27**

**Introduction:**

Sleep disorders are widespread health concerns affecting millions globally. Early detection and management are crucial for improving sleep quality and overall well-being. However, traditional diagnostic methods can be time-consuming and expensive. This project explores the feasibility of using machine learning to predict sleep disorders based on readily available health and lifestyle data.

A key challenge lies in building a model that accurately predicts sleep disorders using a user-friendly interface. This project tackles this challenge by developing a web application that leverages a machine learning model to assess sleep disorder risk and provides personalized well-being recommendations.





**Solution:**

The project utilizes a two-pronged approach:

**Machine Learning Model Development:**

**Data Preprocessing:**

* We employ a sleep disorder dataset (source not disclosed due to confidentiality).
* Data cleaning techniques handle missing values using SimpleImputer ("mean" for features and "most\_frequent" for target variable).
* Categorical features ("Gender", "BMI Category") are encoded with Label Encoder.
* The "Blood Pressure" column is split into "Systolic BP" and "Diastolic BP" for improved model interpretability.

**Machine Learning Models Tested for Sleep Disorder Prediction:**

**1. Decision Tree Classifier:**

* A tree-like structure where each node represents a feature (e.g., age, sleep duration) and a branching decision based on the feature value.
* The decision tree learns a series of rules by splitting the data based on these features until it reaches leaf nodes representing predicted outcomes (sleep disorder or no sleep disorder in this case).
* **Advantages:**
  + Interpretable: The decision tree structure allows for understanding the logic behind its predictions.
  + Relatively fast for training and prediction.
* **Disadvantages:**
  + Prone to overfitting, especially with a large number of features.
  + Decision boundaries can be rigid, potentially leading to less accurate predictions on unseen data.

**2. K-Nearest Neighbors (KNN):**

* Classifies data points based on the similarity to their k nearest neighbors in the training data.
* When a new data point (user input) arrives, the algorithm identifies the k nearest neighbors based on a distance metric (e.g., Euclidean distance).
* The new data point is assigned the majority class label (sleep disorder or no sleep disorder) among its k nearest neighbors.
* **Advantages:**
  + Simple to understand and implement.
  + No explicit training phase required.
* **Disadvantages:**
  + Performance can be sensitive to the choice of k and the distance metric used.
  + Can be computationally expensive for large datasets.

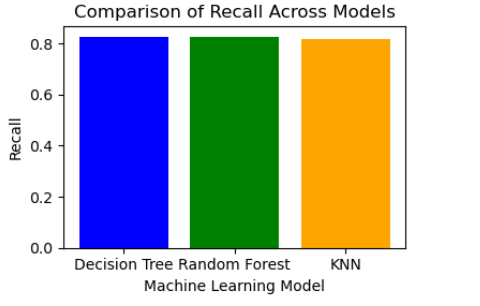
**3. Random Forest Classifier:**

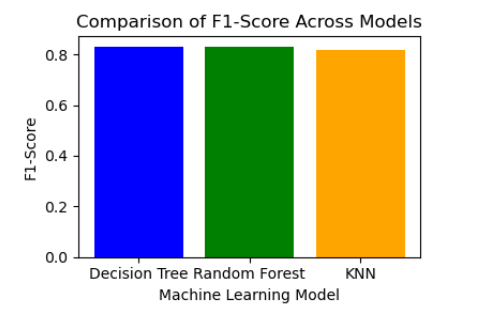
* An ensemble learning method that combines multiple decision trees (referred to as an ensemble).
* Each tree in the forest is trained on a random subset of features and a random sample of the training data with replacement (bootstrapping).
* During prediction, a new data point is passed through all trees in the forest, and the most frequent class label (sleep disorder or no sleep disorder) across all trees becomes the final prediction.
* **Advantages:**
  + Generally more robust to overfitting compared to a single decision tree.
  + Can handle high-dimensional data effectively.
* **Disadvantages:**
  + Less interpretable than a single decision tree due to the ensemble nature.
  + Can be computationally expensive for training.

**Model Selection and Training:**

* Three machine learning algorithms are evaluated: Decision Tree Classifier, KNN, and Random Forest Classifier.
* We choose features based on domain knowledge related to sleep health (age, sleep duration, sleep quality, BMI category, blood pressure, heart rate, daily steps). Feature engineering techniques can be explored further in future work.
* The models are trained on the pre-processed data. Evaluation metrics (precision, recall, F1-score, accuracy) guide the selection of the best performing model.
* Based on the provided information, the Decision Tree Classifier emerges as the leader with an accuracy of **90.67%**.

|  |  |  |  |
| --- | --- | --- | --- |
| Evaluation Metric | Decision Tree | Random Forest | KNN |
| Precision (per class) | [0.7222, 0.9545, 0.8462] | [0.7222, 0.9545, 0.8462] | [0.6667, 0.9535, 0.9091] |
| Precision (micro-averaged) | 0.88 | 0.88 | 0.8667 |
| Precision (macro-averaged) | 0.841 | 0.841 | 0.8431 |
| Precision (weighted-averaged) | 0.8819 | 0.8819 | 0.8828 |
| Recall | 0.8256 | 0.8256 | 0.8178 |
| F1-score | 0.8296 | 0.8296 | 0.817 |





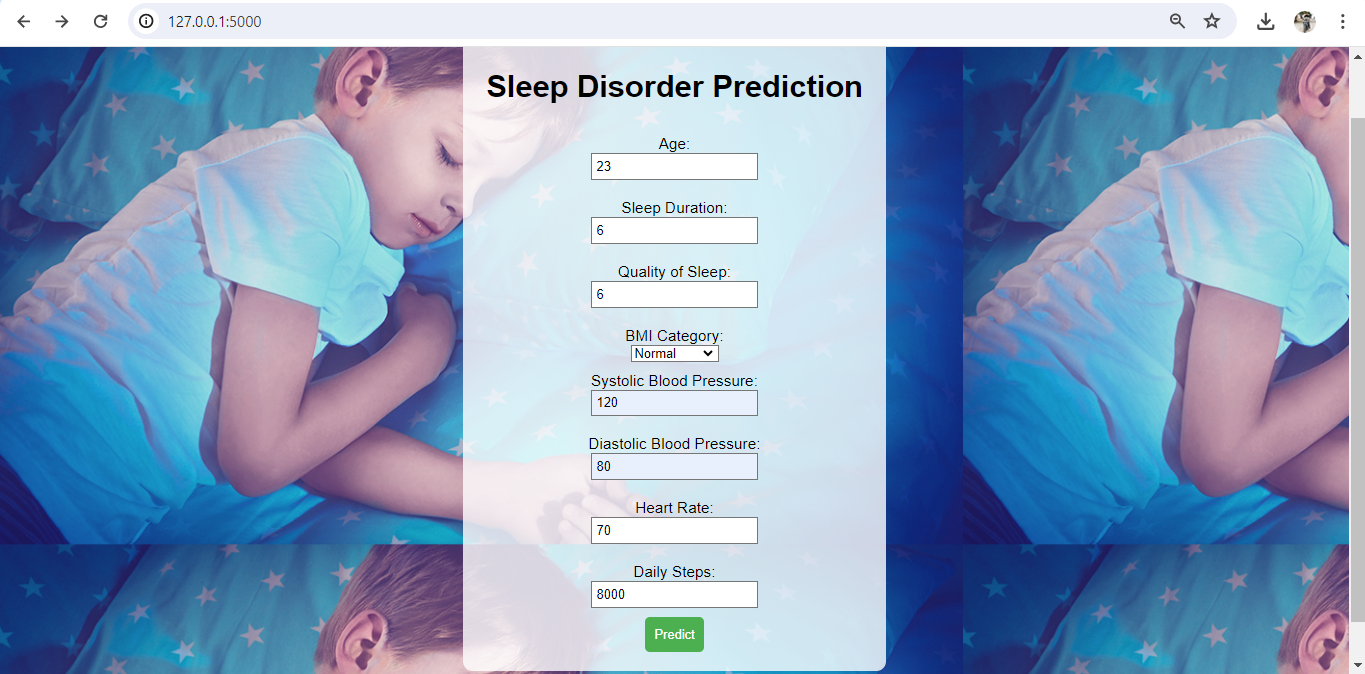
**Web Application Development:**

**Flask Implementation:**

* Flask, a lightweight web framework, is used to create a user-friendly web interface.
* The application allows users to input their health information through a web form.
* The Decision Tree model, loaded at the backend, predicts the risk of sleep disorders based on user input.
* Additionally, the application generates personalized well-being messages based on the user's health metrics. These messages highlight areas for improvement and suggest healthy habits for better sleep and overall health.

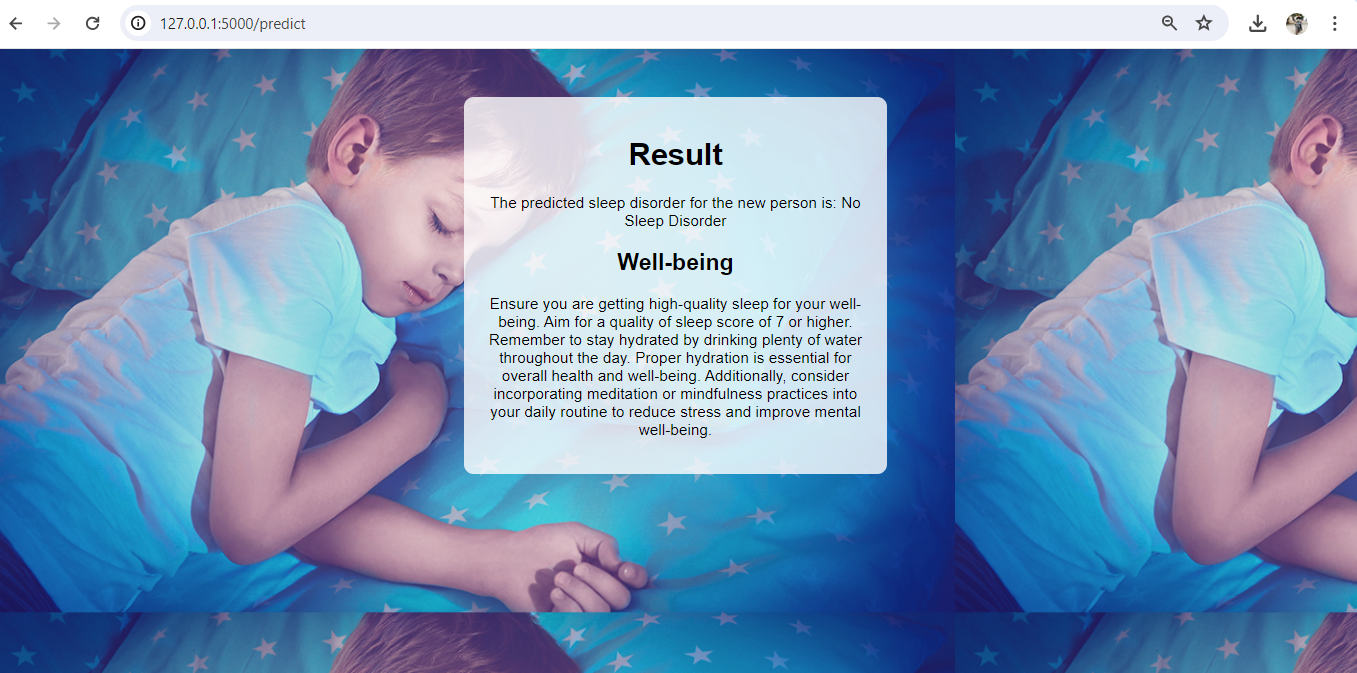
**User Interface:**

* The web application leverages Flask for backend functionality and utilizes HTML and CSS to create a user-friendly interface.
* The user interface allows users to input their health information through a web form, including:
  + Age
  + Sleep Duration (hours)
  + Quality of Sleep (score from 1 to 10)
  + BMI Category (user selection from a dropdown menu)
  + Systolic Blood Pressure (mmHg)
  + Diastolic Blood Pressure (mmHg)
  + Heart Rate (beats per minute)
  + Daily Steps
* The user can submit the form to receive a prediction and personalized well-being recommendations.



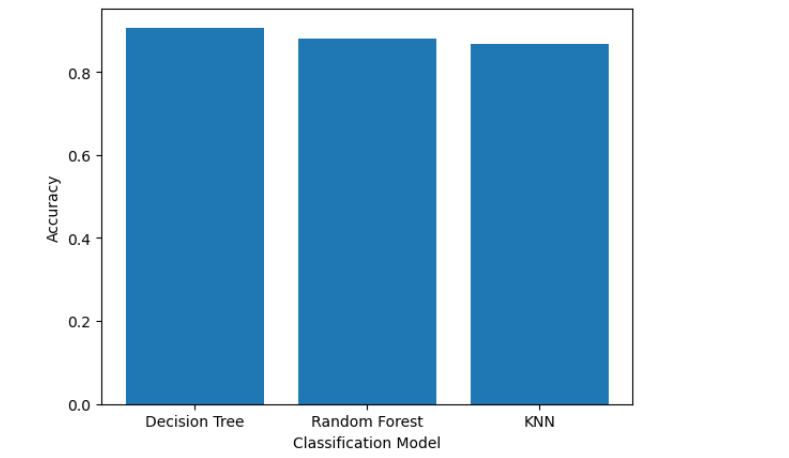
**Backend Functionality:**

* When the user submits the form, the Flask application retrieves the user's input data.
* The retrieved data is preprocessed (e.g., converting categorical data using the trained LabelEncoder) to match the format the Decision Tree model expects.
* The preprocessed data is used as input for the pre-trained Decision Tree model.
* The model predicts the risk of sleep disorders based on the user's input.
* The application generates a personalized well-being message based on the user's health metrics. This message incorporates:
  + Sleep disorder prediction result.
  + Feedback on specific health metrics that might require improvement for better sleep and overall health.
  + Additional recommendations for well-being (e.g., increasing physical activity, reducing caffeine intake, staying hydrated).



**Empirical Experiments:**

The project employs three machine learning models (Decision Tree, KNN, Random Forest) as the core experiment. The Decision Tree Classifier achieves the highest accuracy (**90.67%**) compared to Random Forest (**88.00%**) and KNN (**86.67%**). This suggests that the Decision Tree model effectively learns the patterns within the sleep disorder dataset for risk prediction.



**Discussion:**

While the Decision Tree Classifier demonstrates promising results, some limitations are worth considering:

* **Data Dependence:** The model's accuracy is highly dependent on the quality and representativeness of the training data. A more diverse and extensive dataset could potentially improve generalizability.
* **Model Complexity:** Decision Trees can be prone to overfitting, especially with a large number of features. Implementing pruning techniques or exploring hyperparameter tuning could mitigate this limitation.
* **Interpretability vs. Performance:** Decision Trees offer good interpretability, making it easier to understand the factors influencing the model's predictions. However, more complex models like Random Forests might achieve higher accuracy at the cost of interpretability.

Future exploration could involve:

* **Feature Engineering:** Exploring feature engineering techniques to potentially create new features that improve model performance.
* **Model Ensembling:** Investigating ensemble methods like Random Forests for potentially higher accuracy while maintaining interpretability.
* **Integration with Wearables:** Integrating the application with wearable sleep trackers for real-time data collection and personalized sleep hygiene recommendations.

**Conclusion:**

This project demonstrates the potential of using a Decision Tree Classifier within a user-friendly web application to predict sleep disorders and provide personalized well-being recommendations. The application can be a valuable tool for raising sleep health awareness and encouraging healthy lifestyle habits. By addressing the limitations and exploring future directions, this project can contribute to the development of more robust and informative sleep disorder prediction systems.