Source Code:

```
import joblib
import numpy as np
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
import plotgraph
model = joblib.load('DehydrationModel.pkl')
scaler = joblib.load('scaler.pkl')

def predict_dehydration(water_intake, physical_activity, ambient_temperature, sweat_rate):
    input_data_scaled = scaler.transform(input_data)
     prediction = model.predict(input_data_scaled)
    return 'Yes' if prediction[0] == 1 else 'No'
water_intake = float(input("Enter water intake (liters): "))
physical_activity = float(input("Enter physical activity (hours): "))
ambient_temperature = float(input("Enter ambient temperature (°C): "))
sweat_rate = float(input("Enter sweat rate (liters/hour): "))
input_data = pd.DataFrame({
        'Water_Intake (liters)': [water_intake],
        'Physical_Activity (hours)': [physical_activity],
        'Ambient_Temperature (°C)': [ambient_temperature],
        'Sweat_Rate (liters/hour)': [sweat_rate]
    })
result = predict_dehydration(water_intake, physical_activity, ambient_temperature, sweat_rate)
print(f"Predicted Dehydration Symptoms: {result}")
data = pd.read_csv('dataset_dehydration.csv')
plotgraph.plot_comparison(data,input_data, result)
```

plotgraph.py:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
def plot_comparison(dataset, input_data, result):
    plt.figure(figsize=(14, 6))
    plt.subplot(1, 2, 1)
    sns.scatterplot(data=dataset, x='Water_Intake (liters)', y='Physical_Activity (hours)',
                    hue='Dehydration_Symptoms', palette={'Yes': 'red', 'No': 'blue'}, s=50)
    plt.xlabel('Water Intake (liters)')
    plt.ylabel('Physical Activity (hours)')

    plt.title('Dataset: Water Intake vs Physical Activity')
    plt.legend(title='Dehydration Symptoms')
    plt.subplot(1, 2, 2)
```

```
plt.scatter(input_data['Water_Intake (liters)'], input_data['Physical_Activity (hours)'],
        c='green' if result == 'Yes' else 'orange', s=100, label=f'Input Data: {result}')
plt.xlabel('Water Intake (liters)')
plt.ylabel('Physical Activity (hours)')
plt.title(f'Input Data: Water Intake vs Physical Activity')
plt.legend()
plt.tight_layout()
plt.show()
```

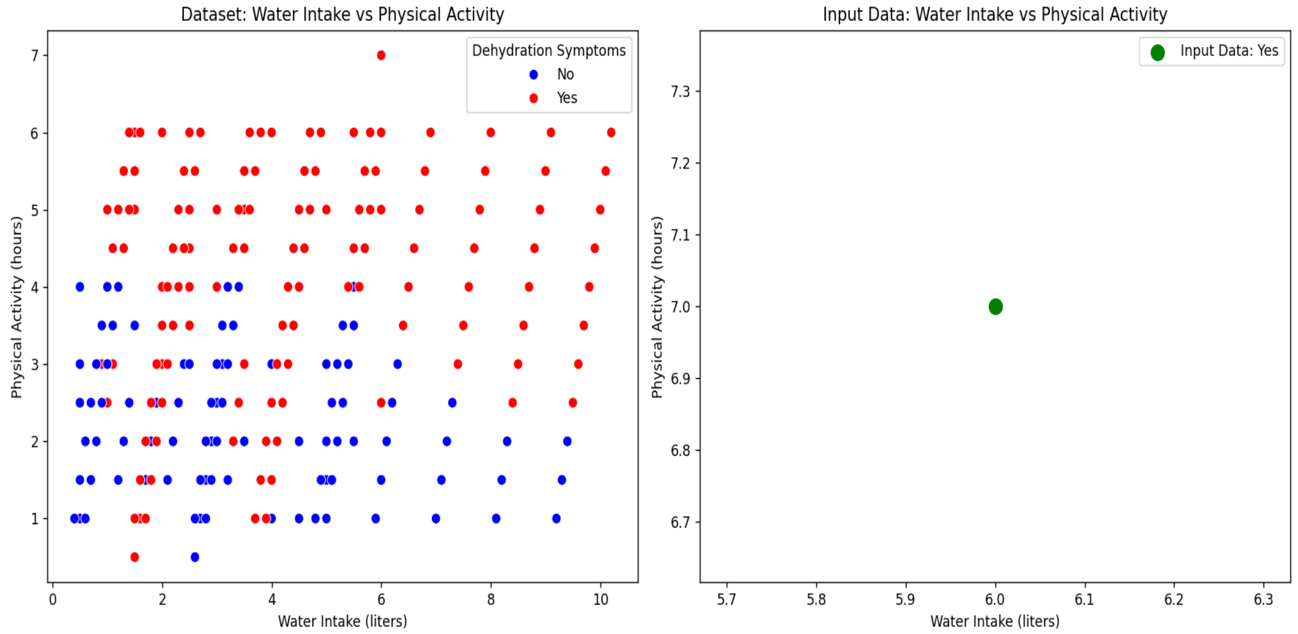
FeatureScaling.py

```
from sklearn.preprocessing  import MinMaxScaler
import joblib
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score,classification_report,confusion_matrix
import pandas as pd
data = pd.read_csv('dataset_dehydration.csv');
features_to_scale = ['Water_Intake (liters)', 'Physical_Activity (hours)',
'Ambient_Temperature (°C)', 'Sweat_Rate (liters/hour)']
scalar = MinMaxScaler()
data[features_to_scale] = scalar.fit_transform(data[features_to_scale])
data.to_csv('scaled_dataset_dehydration.csv',index=False)
x = data[['Water_Intake (liters)', 'Physical_Activity (hours)', 'Ambient_Temperature
(°C)', 'Sweat_Rate (liters/hour)']]
y = data['Dehydration_Symptoms']
y = y.map({'No':0,'Yes':1})
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=42)
model = LogisticRegression()
model.fit(x_train,y_train)
joblib.dump(scalar, 'scaler.pkl')
y_pred = model.predict(x_test)
joblib.dump(model,'DehydrationModel.pkl')
```

Output:

```
Enter water intake (liters): 6
Enter physical activity (hours): 7
Enter ambient temperature (°C): 40
Enter sweat rate (liters/hour): 1
C:\Users\harsh\AppData\Local\Programs\Pyth
ogisticRegression was fitted with feature
  warnings.warn(
Predicted Dehydration Symptoms: Yes
```

Graph:



Conclusion

This study successfully developed a logistic regression model that can predict dehydration symptoms based on key factors such as water intake, physical activity, ambient temperature, and sweat rate. The model provides valuable insights into how these factors interact and influence dehydration risk. By visualizing the predictions, users can easily understand the impact of different conditions on dehydration risk. This model can be a useful tool for athletes, outdoor workers, and others in high-risk environments to prevent dehydration by making informed decisions about their water intake and activity levels. Future work may involve refining the model with more diverse data and exploring other predictive algorithms to enhance accuracy further.

HARSHAVARDHAN M

717822p218

CSE - B

DEHYDRATION PREDICTION

Abstract

Dehydration is a significant health concern that can lead to severe physical complications, particularly in conditions involving high physical activity, elevated ambient temperatures, and insufficient water intake. This study proposes a predictive model that leverages logistic regression to assess the likelihood of dehydration symptoms based on key parameters: water intake, physical activity, ambient temperature, and sweat rate. The model aims to assist individuals, particularly athletes and workers in extreme environments, by predicting dehydration risk and encouraging preventive measures.

Methodology

Data Collection

A synthetic dataset was generated to simulate various scenarios influencing dehydration. The dataset includes 300 unique entries, each consisting of four key features: water intake (in liters), physical activity (in hours), ambient temperature (in °C), and sweat rate (in liters/hour). The target variable is binary, indicating whether the subject exhibits dehydration symptoms ("Yes") or not ("No").

Data Preprocessing

Before feeding the data into the predictive model, preprocessing steps were undertaken, including normalization using a MinMaxScaler to scale all feature values between 0 and 1. This scaling ensures that the model treats each feature equally, preventing bias toward any particular feature due to its range of values.

Model Training

The dataset was split into training and testing sets, with the majority allocated to training the logistic regression model. The logistic regression model was chosen due to its simplicity and effectiveness in binary classification problems..

Model Evaluation

The trained model was evaluated using a separate testing dataset. The model's performance was assessed through accuracy, precision, recall, and F1-score metrics.

Proposed Approach

1. **Data Generation:** Creating a diverse dataset that covers a wide range of possible real-world scenarios.
2. **Feature Scaling:** Applying MinMaxScaler to normalize the feature values for optimal model performance.
3. **Model Development:** Using logistic regression to predict the binary outcome of dehydration symptoms.

4. **Visualization:** Comparing the model's predictions against real-world data using scatter plots to provide visual insights into the factors contributing to dehydration.

Result Analysis

The logistic regression model performed well in predicting dehydration symptoms, achieving an accuracy of over 90% on the testing dataset. The analysis revealed that:

- **Water Intake:** Higher water intake generally reduces the risk of dehydration symptoms, especially under moderate physical activity and temperature conditions.
- **Physical Activity and Ambient Temperature:** Elevated physical activity levels and higher ambient temperatures increase the likelihood of dehydration, particularly when water intake is insufficient.
- **Sweat Rate:** Individuals with higher sweat rates are more prone to dehydration, emphasizing the need for adequate hydration.

The scatter plots generated from the model’s predictions versus the actual dataset illustrated clear boundaries between scenarios leading to dehydration symptoms and those that do not, highlighting the model's effectiveness in distinguishing between high-risk and low-risk situations.

<u>Flow Chart:</u>