

Personalized Wellness Insights

1. Introduction

- **Objective:** This project aims to analyze and cluster wellness sessions based on metrics such as accuracy of poses, duration, and engagement levels to generate personalized insights for improving overall wellness experiences.
- **Background:** Wellness sessions, such as yoga or fitness routines, involve multiple dimensions of performance. Personalizing insights based on individual session data can enhance user experience, motivation, and outcomes. This project applies data analysis, clustering, and API deployment to create a streamlined system for these insights.

2. Approach

- **Overall Workflow:** The following steps were taken to build the system:
 - Simulating wellness session data.
 - Preprocessing the data for analysis.
 - Applying clustering to segment sessions based on similar attributes.
 - Visualizing the clustered data for interpretability.
 - Deploying a Flask-based API to serve insights.
- **Tools Used:**
 - Python (Pandas, NumPy, Matplotlib, Seaborn)
 - Scikit-learn (Clustering, Metrics)
 - Flask (API Development)

3. Data Preprocessing

- **Dataset:**
 1. Simulated data consists of 50 sessions with fields:
 - PoseAccuracy: Pose accuracy in percentage (70–100%).
 - Duration: Session duration in minutes (20–60).
 - Engagement: Engagement score on a scale of 0.5 to 1.0.
 - Consistency: A derived feature calculated as $\text{PoseAccuracy} * \text{Engagement}$.
- **Preprocessing Steps:**
 1. **Data Simulation:** Randomized data generation using NumPy.
 2. **Feature Scaling:**
 - Used MinMaxScaler to scale PoseAccuracy, Duration, and Engagement to a uniform range for clustering.

3. Visualization:

- Bar plots and scatter plots used to understand the distribution of metrics across sessions.

Example snippet of preprocessed data:

SessionID	PoseAccuracy	Duration	Engagement	Consistency	Cluster
1	81.23620357	37	0.7806217	63.4147444	0
2	98.52142919	45	0.6914634	68.1239661	0
3	91.95981825	53	0.985856	90.659143	1
4	87.95975453	29	0.9244569	81.3150031	1
5	74.68055921	55	0.8608648	64.2898617	0

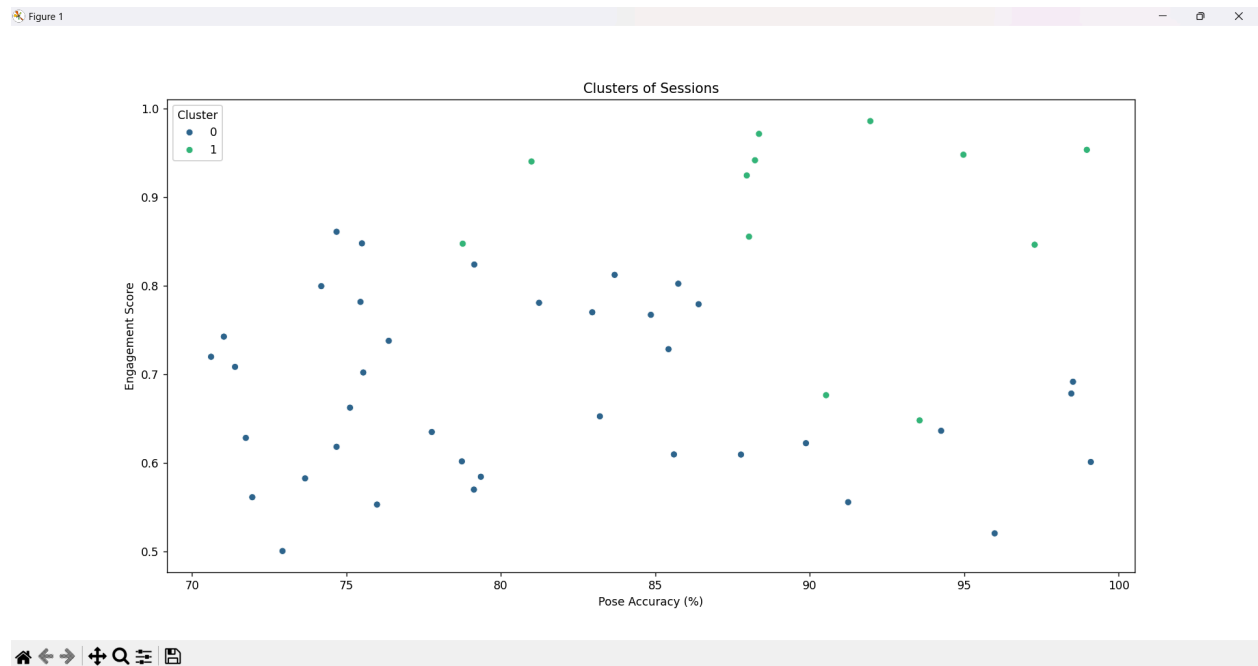
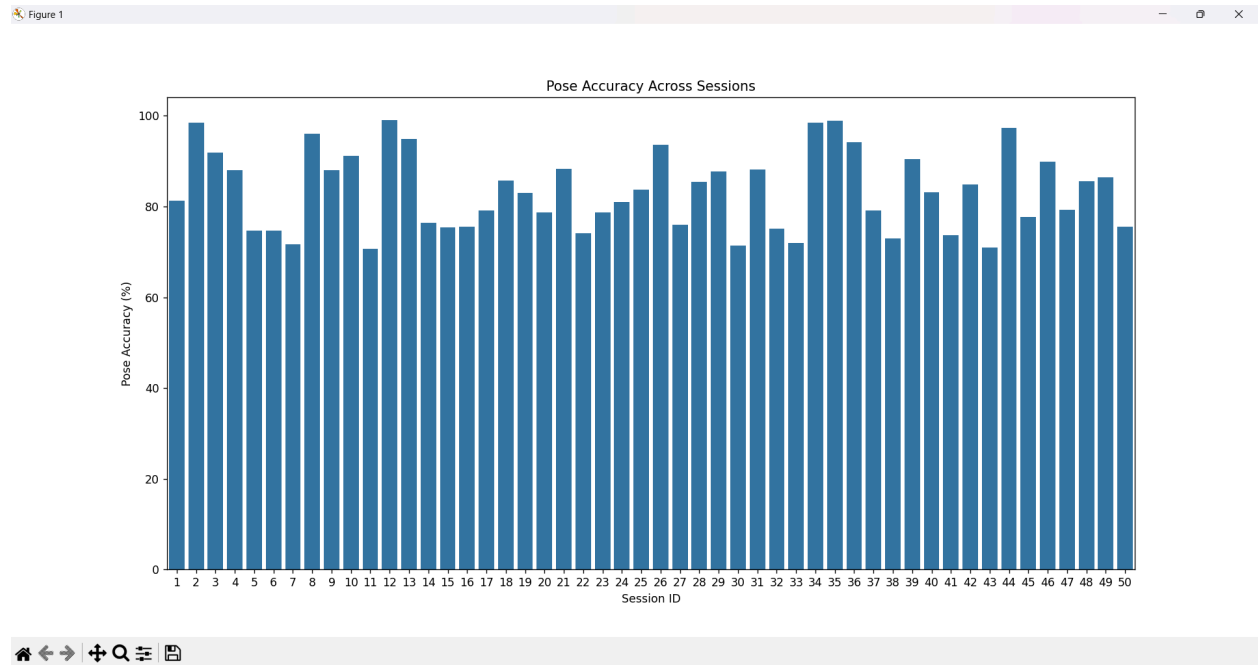
4. Model Architecture

- **Clustering Algorithm:**
 - Chose **K-Means Clustering** for its simplicity and ability to partition the data into well-defined clusters.
- **Number of Clusters:**
 - Selected 2 clusters to group sessions into high-performance and moderate-performance categories.
- **Evaluation:**
 - Used **Silhouette Score** to assess cluster quality. The calculated score was 0.6, indicating moderately good clustering.
- **Hyperparameters:**
 - n_clusters=2 and random_state=42 for reproducibility.
- **Visualization:**
 - Plotted clusters to observe grouping based on PoseAccuracy and Engagement.

5. Results

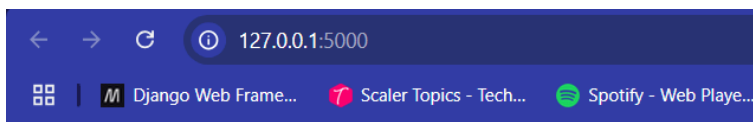
- **Cluster Analysis:**
 - **Cluster 0:** Represents sessions with higher pose accuracy and engagement.
 - **Cluster 1:** Represents sessions with moderate pose accuracy and engagement.
- **Key Metrics:**
 - Silhouette Score: 0.6
 - Data Distribution: Around 30% of sessions fall into Cluster 0, and 70% into Cluster 1.
- **Visual Representation:**

- Include scatter plots and bar charts showcasing clusters.



6. Flask API

- **Endpoints:**
 - `/`: Displays a welcome message (*“Welcome to the Personalized Yoga Sessions Insights API!”*).
 - `/insights`: Provides the session data with cluster assignments in JSON format.
- **API Functionality:**
 - Hosts the clustered data as a REST API for integration with other applications.
- **Deployment:**
 - Deployed locally for testing and can be scaled to cloud platforms (e.g., AWS, Azure) for broader accessibility.



7. Conclusion

- The project demonstrates how clustering wellness session data can uncover actionable insights. By scaling this system to real-world data and applications, it can significantly enhance user experiences in wellness programs.