Healthcare Dataset Analysis

In this project, we performed an analysis on a healthcare dataset with the goal of predicting medical conditions based on various categorical features. The steps and results of our analysis are outlined below.

Data Preprocessing

Before training our model, we preprocessed the data as follows:

1. Ordinal Encoding: Converted categorical features into numerical values using ordinal encoding.

2. Data Splitting: Split the dataset into training and testing sets to evaluate model performance.

Model Training

We employed a Random Forest Classifier to predict medical conditions from the given features. To optimize the model, we used RandomizedSearchCV , a technique that performs random combinations of hyperparameters and selects the best combination based on cross-validated performance.

Model Evaluation

The performance of the Random Forest model was evaluated using the following key metrics:

- Accuracy: 0.79

- Precision: 0.81

- Recall: 0.79

- F1 Score: 0.79

These results indicate that the Random Forest model is fairly effective at predicting medical conditions from the given features, demonstrating a balanced performance across different evaluation metrics.

Conclusion

The Random Forest model showed a balanced and satisfactory performance in predicting medical conditions based on the dataset. The results suggest that the model can be a reliable tool for medical predictions, with an accuracy of 79% and consistent precision, recall, and F1 score.

For detailed code and analysis, please refer to the accompanying files in this repository.

Name :- Pranav Parasar
E-mail:- pranavparasar99@gmail.com