

# Exploring the Effectiveness of PCA in Predictive Analysis of Sleep Disorders

## ABSTRACT

This study investigates the predictive analysis of sleep disorders using advanced machine learning and Principal Component Analysis (PCA) to manage high-dimensional data. Addressing common issues like overfitting and inefficiency, PCA is employed to refine the "Sleep Health and Lifestyle" dataset, optimizing the precision of models such as RandomForest, Logistic Regression, SVC, and KNN.

## INTRODUCTION

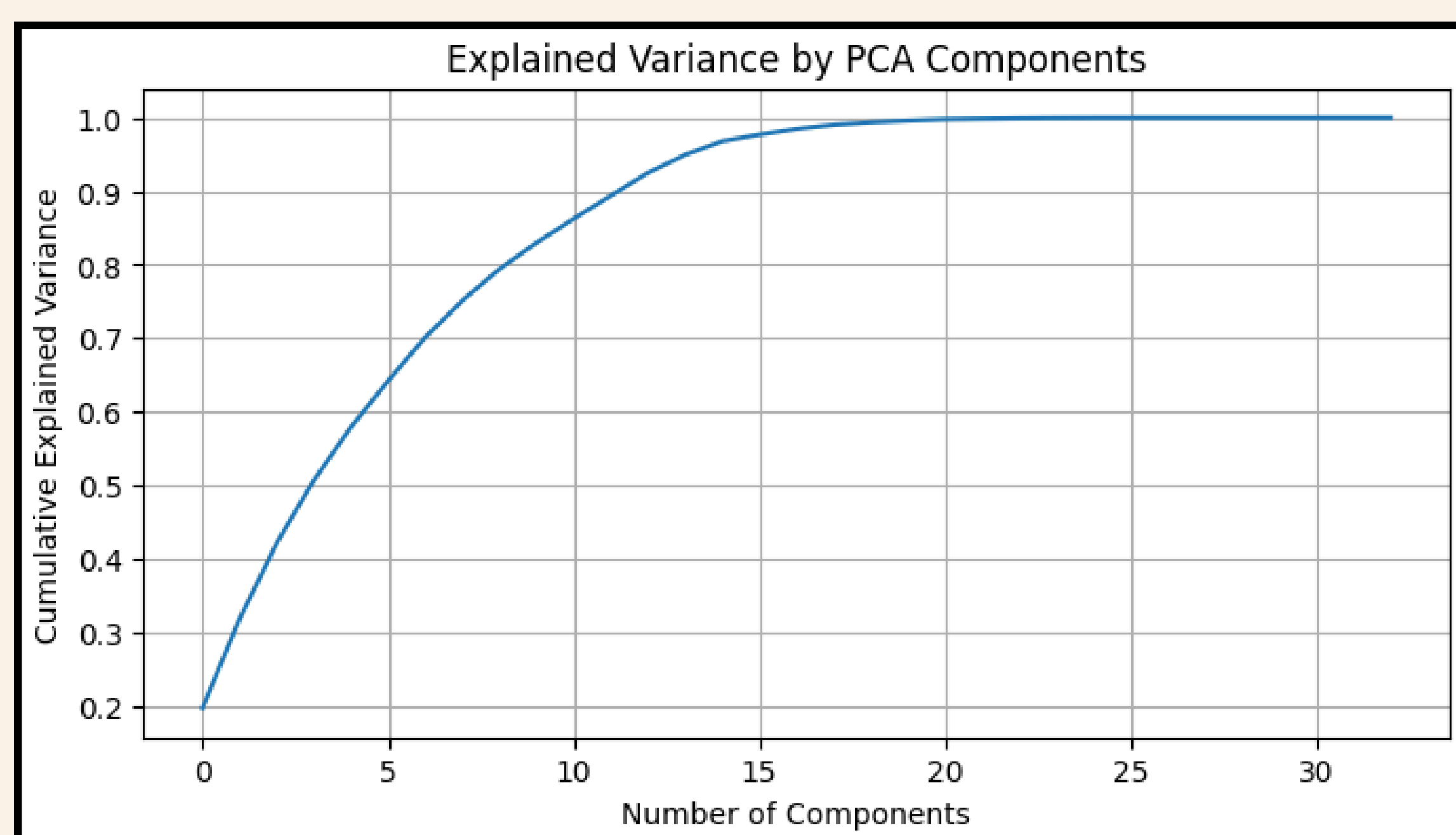
This research utilizes machine learning and PCA to predict sleep disorders, addressing the complexity of high-dimensional data. By using the "Sleep Health and Lifestyle" dataset, we improve predictive modeling with PCA, boosting accuracy without sacrificing essential data integrity. Our method paves the way for more precise and understandable healthcare analytics.

## METHODOLOGY

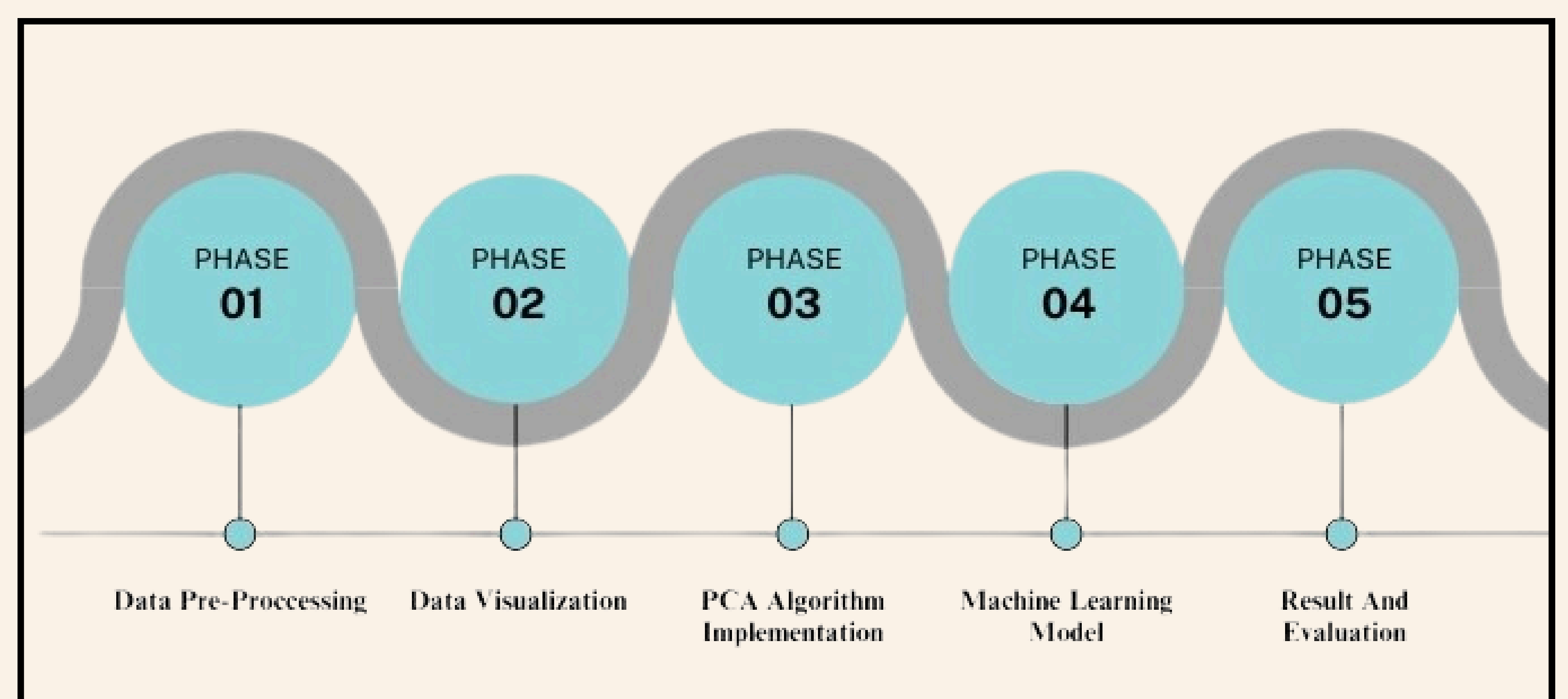
The research methodology proceeds through a five-phase process:

- (1) Data pre-processing for data normalization
- (2) Visualization to explore data characteristics
- (3) PCA for dimensionality reduction
- (4) Machine learning model construction and training
- (5) Evaluation of the results to assess the model's predictive performance.

## PROJECT DESCRIPTION



PCA Implementation Process



Sequential Flowchart of the Analytical Process

## RESULTS

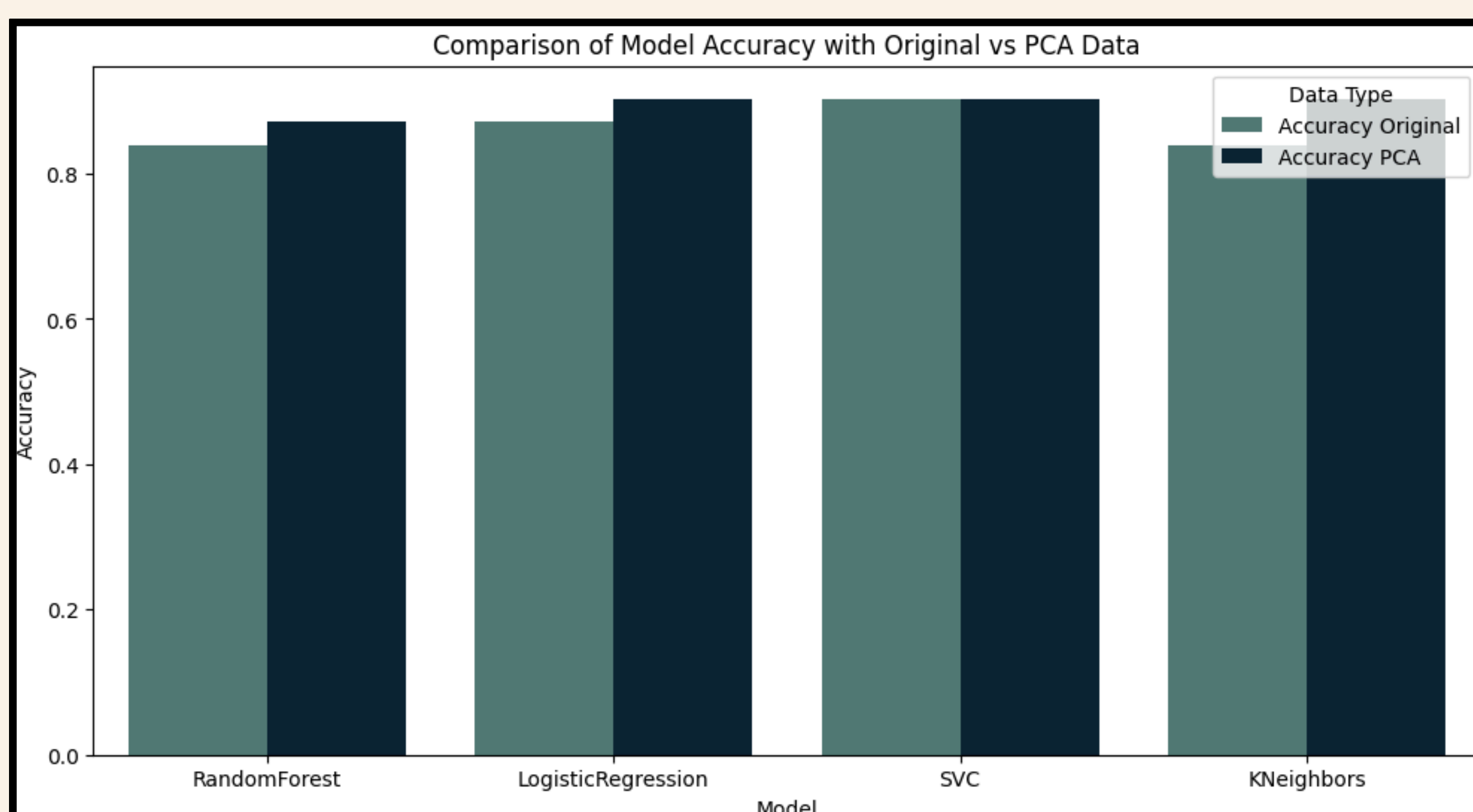


Table 2 Model Performance Before and After PCA

Model	Accuracy Before PCA	Accuracy After PCA	Time Before PCA	Time After PCA
RandomForest	83.87%	93.54%	0.177907	0.168384
LogisticRegression	87.10%	90.32%	0.005161	0.002880
SVC	90.32%	90.32%	0.001254	0.001681
KNeighbors	83.87%	90.32%	0.000565	0.00670

Table 3 Performance Comparison Table

Model	Our Model Accuracy	Reported Accuracy	Difference
RandomForest	93.54%	92.15%	+1.39%
Support Vector Machine	90.32%	90.02%	+0.30%
k-Nearest Neighbors	90.32%	86.91%	+3.41

## CONCLUSION

This study confirms PCA as a pivotal tool in enhancing machine learning for sleep disorder prediction, improving accuracy for certain models and underscoring the importance of tailored dimensionality reduction in healthcare analytics.

## TEAM MEMBERS

SARAH ALSHUMAYRI - REEMA ABDALLAH - LEEN SHARAB  
COLLEGE OF ENGINEERING  
COMPUTER SCIENCE DEPARTMENT  
SUPERVISOR: P. PASSENT M. ELKAFAWY