Title: Predicting Hypertension Risk- A Comparative Study of Machine Learning Regression Techniques

Background:
Hypertension is a leading cause of death worldwide. Early identification of individuals at high risk of developing hypertension is crucial for timely intervention and prevention.

Machine learning techniques have shown promise in predicting hypertension risk based on various health metrics and risk factors. This study aims to explore and compare the performance of different machine learning approaches in predicting hypertension risk using a dataset of patient health information.

Problem Statement:
The goal of this study is to develop an accurate and efficient model for predicting hypertension risk based on patient data, including demographics, lifestyle factors, and health metrics. By comparing various machine learning techniques, we aim to identify the most effective approach for this task. The study will address the following research questions:

1. Which machine learning technique (K-Means Clustering, PCA, linear regression, ridge regression, lasso regression, SFS) yields the highest accuracy in predicting hypertension risk?

2. What are the most important features or risk factors contributing to hypertension risk prediction?

3. How can we optimize the selected machine learning model to improve its performance and generalizability?

Methodology:
The study will follow these steps:

1. Data preprocessing: Clean and preprocess the dataset, handling missing values, outliers, and categorical variables. Perform feature scaling and normalization as needed.

2. Exploratory data analysis: Conduct descriptive statistics and visualizations to gain insights into the distribution of variables and their relationships with hypertension risk.

3. Feature selection: Apply techniques such as PCA and SFS to identify the most relevant features for hypertension risk prediction.

4. Model training and evaluation: Implement and train various machine learning models, including K-Means Clustering, Linear Regression, Ridge regression, Lasso regression. Use appropriate evaluation metrics (e.g., MSE, RMSE) to assess model performance. Employ cross-validation to ensure robustness.

5. Model comparison: Compare the performance of different models and select the best-performing approach. Analyze the strengths and limitations of each technique.

6. Interpretation and insights: Interpret the results, identifying the most important risk factors and their impact on hypertension risk. Discuss the implications of the findings for clinical decision-making and patient care.