**Introduction :**

This report summarizes the findings from Recursive Feature Elimination (RFE) applied to the **Diabetes dataset**.  
RFE was used to identify the most relevant features affecting **diabetes progression**.  
The analysis includes feature ranking, comparison with other selection methods, **and key dataset insights**.

**Feature Coefficients at Each RFE Iteration:**

The table below shows how feature importance changed as features were eliminated step by step

A black and white screen with numbers

AI-generated content may be incorrect.

Top 3 Most Important Features:

A black screen with white text

AI-generated content may be incorrect.

What we can infere is

* s1 has the strongest negative impact , Higher values slow down diabetes progression.
* s5 is highly positive , Higher values increase diabetes progression significantly.
* bmi (Body Mass Index) is a major contributor , Obesity is a strong factor in diabetes risk.

Comparison of Initial vs. Final Feature Selection :

A screen shot of a computer

AI-generated content may be incorrect.

* RFE retained all 10 features because removing any feature led to an R² drop greater than 0.01.
* The strongest predictors (s1, s5, and bmi) remained highly ranked even after RFE.
* Age and sex had the lowest impact but were still retained since they contributed meaningfully.

Conclusion :

RFE retained all 10 features because removing any one caused an R² drop greater than 0.01, meaning each feature contributed enough to keep. BMI, s5, and s1 were the strongest predictors, showing that body mass and blood serum levels are key factors in diabetes progression. BP had moderate importance, while age and sex had the least impact but were still retained as they added some value.

Unlike LASSO, which might have forced some features to zero, RFE ranked features without removing any, ensuring all useful predictors stayed. Diabetes progression is influenced by multiple factors working together, not just one, making it important to keep all relevant features for better predictions.

This graph shows the changes in r2 when number features are retained

A graph with a line

AI-generated content may be incorrect.

Key Findings of mine :

The R² score, which measures how well the model explains the variability in diabetes progression, was 0.4526 when using all 10 features, indicating that the model explains 45.26% of the variance. However, using the 0.01 R² improvement threshold, RFE determined that only 6 features were necessary, as removing any more led to a significant drop in performance. This suggests that while all features contribute to some extent, selecting the most relevant ones improves model efficiency without sacrificing accuracy. By reducing the number of features from 10 to 6, we maintain a strong predictive ability while ensuring a more interpretable and optimized model.