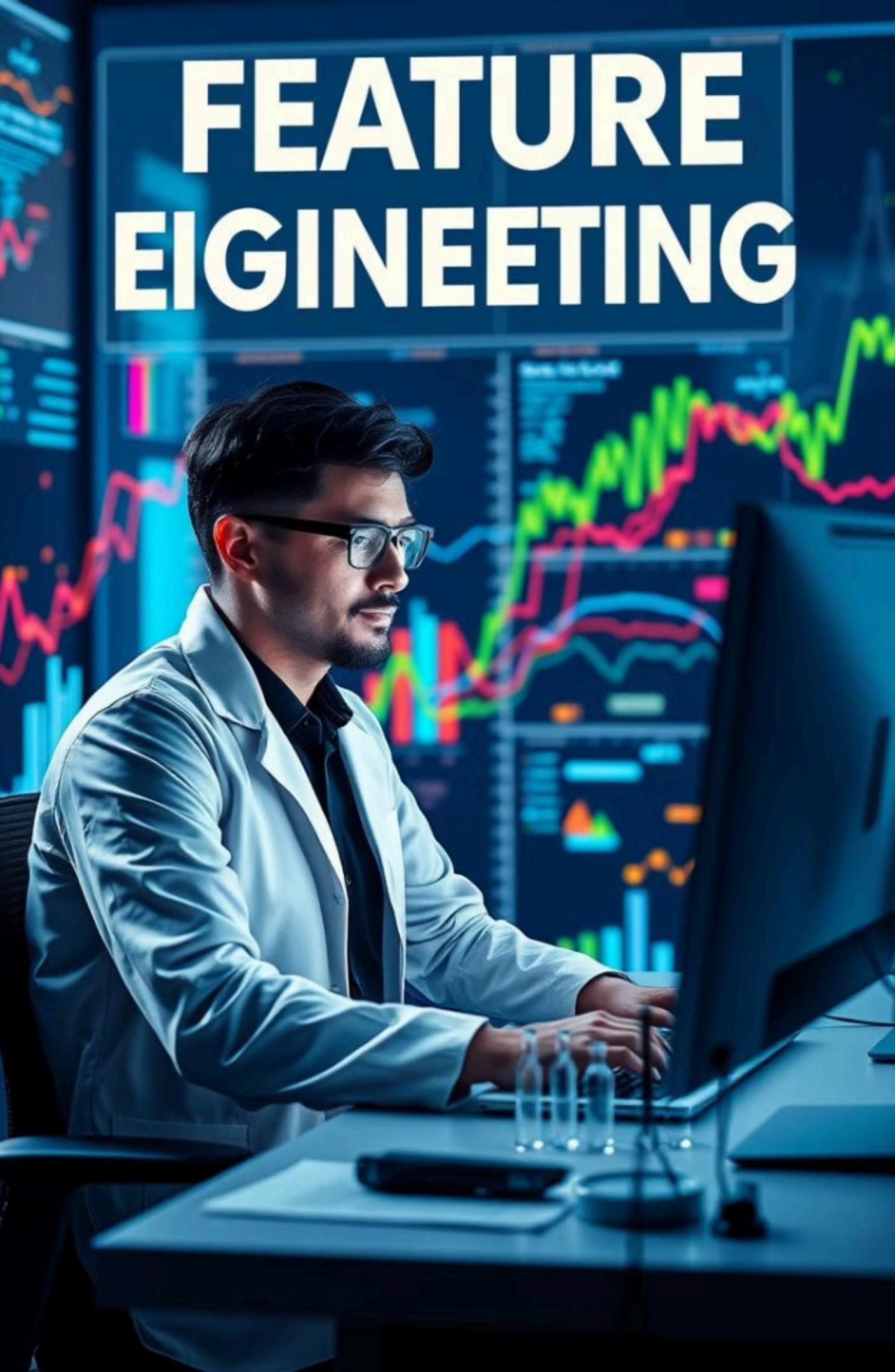


A complex network graph composed of numerous small, glowing blue and white dots connected by thin white lines. A few larger, more prominent glowing blue and white dots are scattered throughout, some with radiating lines. The overall effect is a futuristic, digital representation of data connections.

Feature Engineering & Selection Techniques

Data transformation and selection are key for effective machine learning.

by Aravind K



Introduction

Feature Engineering: Transform raw data into meaningful features.

Feature Selection: Choose most relevant features to improve model accuracy.

1

Improved Performance

Enhanced accuracy and reduced overfitting.

2

Reduced Complexity

Simpler models and faster training times.

3

Better Insights

Understanding data and identifying important variables.

Dataset Overview

California Housing Dataset: Predict median house values.

Includes features like population, income, and age.

Feature	Description
MedHouseVal	Median house value in units of 100,000
Population	Population of the block group
AveOccup	Average number of rooms per household
Latitude	Latitude of the block group
Longitude	Longitude of the block group

California House Value

for a car the clustered meals





Feature Engineering Techniques

Transforming raw data into meaningful features.

- 1 Encoding Categorical Data**
Convert categorical features into numerical representations.
- 2 Scaling Numerical Features**
Normalize data to a common scale.
- 3 Interaction Features**
Create new features by combining existing features.
- 4 Polynomial Features**
Generate higher-order terms from existing features.

Feature Selection Techniques

Selecting most relevant features to improve model performance.

Variance Thresholding

Remove features with low variance.

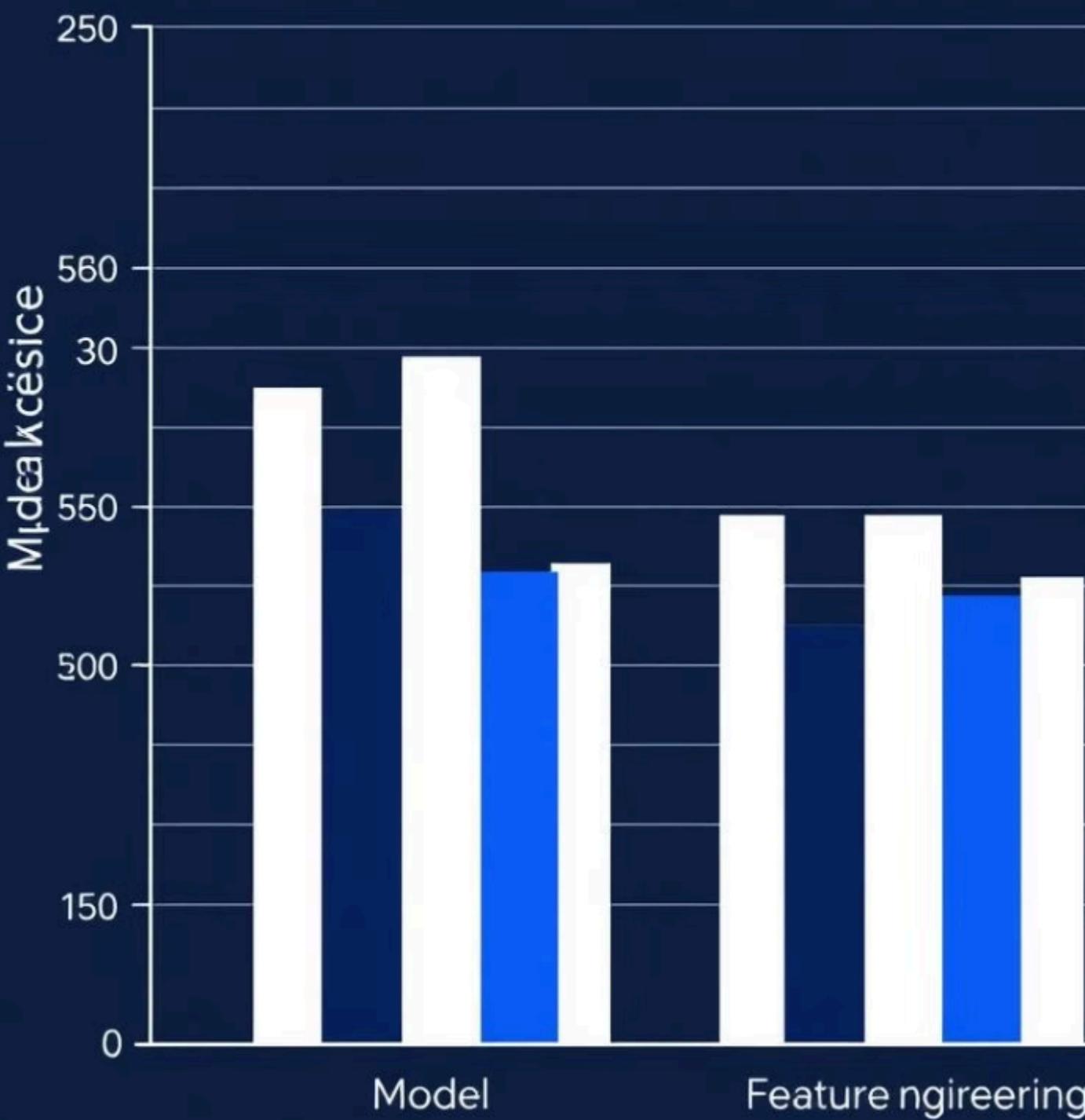
Univariate Selection

Select features based on statistical tests.

Recursive Feature Elimination

Iteratively remove least important features.

Pevelve tegiretlecing
tour have **festtiireaunireniing**
rofless call bd feature rugineet wiithem



Results

After applying feature engineering and selection techniques:

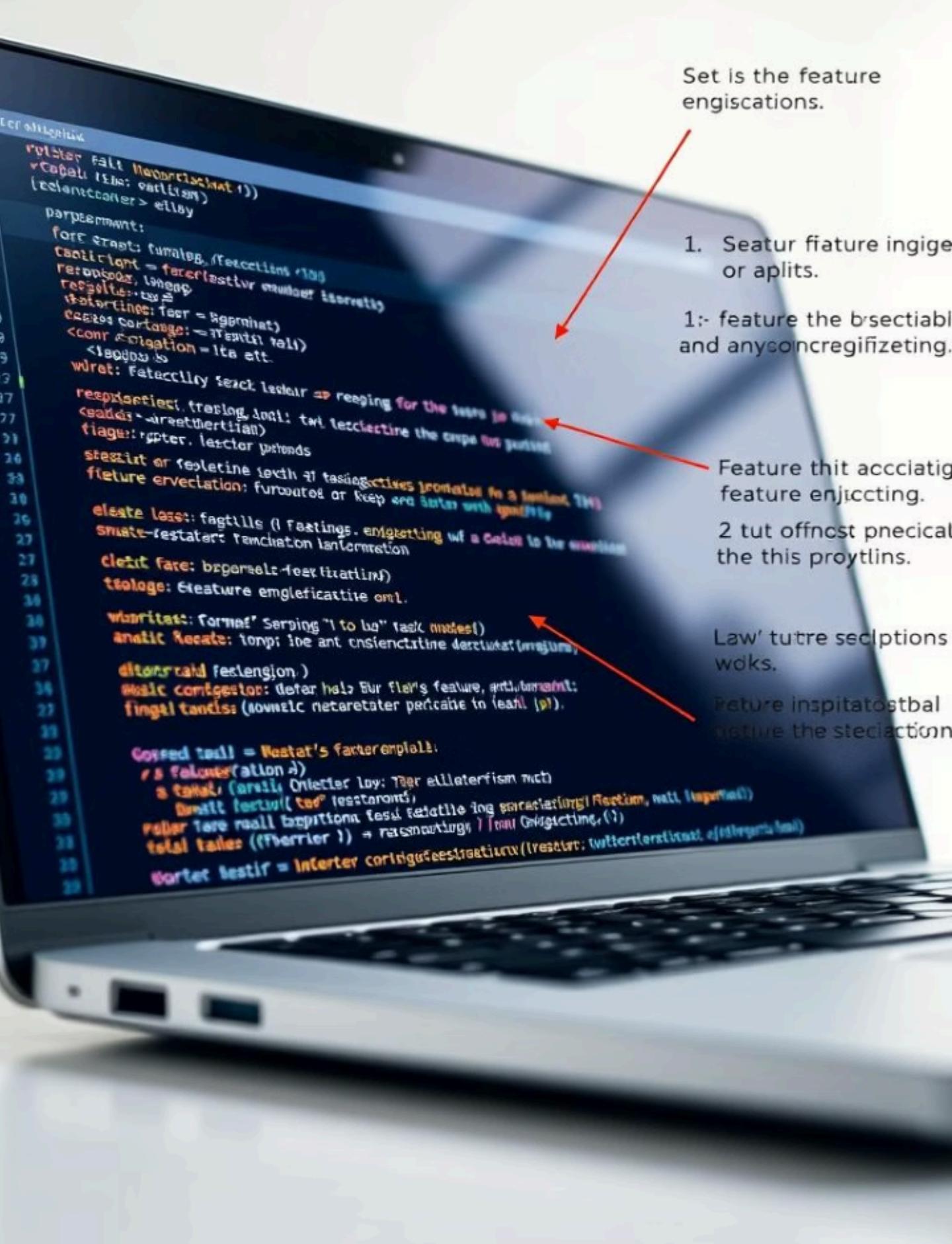
- 1 Variance Thresholding
Reduced features from X to Y.
- 2 SelectKBest
Top 5 features: Average Income, House Age, etc.
- 3 RFE
Retained features based on importance scores.

A portrait of a young man with dark hair and a beard, wearing black-rimmed glasses and a white lab coat over a blue shirt. He is smiling and holding a white coffee cup in his right hand and a dark laptop in his left hand. The background is a solid teal color.

Conclusion

Feature engineering and selection significantly improve model performance.

Essential steps for efficient and accurate machine learning workflows.



References

California Housing Dataset: scikit-learn.

Tools: Python, Jupyter Notebook, Scikit-learn, Pandas.