Feature Selection across Models

Ryan Hsu
CS 677 - Final Project
4/27/2022

Agenda

- Problem Description
- Feature Selection and Models
- Dataset
- Python Code Tour
- Findings

Problem Description

This project attempts to dive into the effect feature selection makes on different prediction models.

To visualize this, this project attempts to find the most important features that lead to a student doing well in an academic course.

Feature Selection and Models

Feature Selection

- Select K Best Features
- RFE Feature Selection

Models for Classification

- Naive Bayesian,
- Decision Tree
- k-NN
- Polynomial SVM

Dataset

Higher Education Students Performance Evaluation Dataset Data Set

Data:

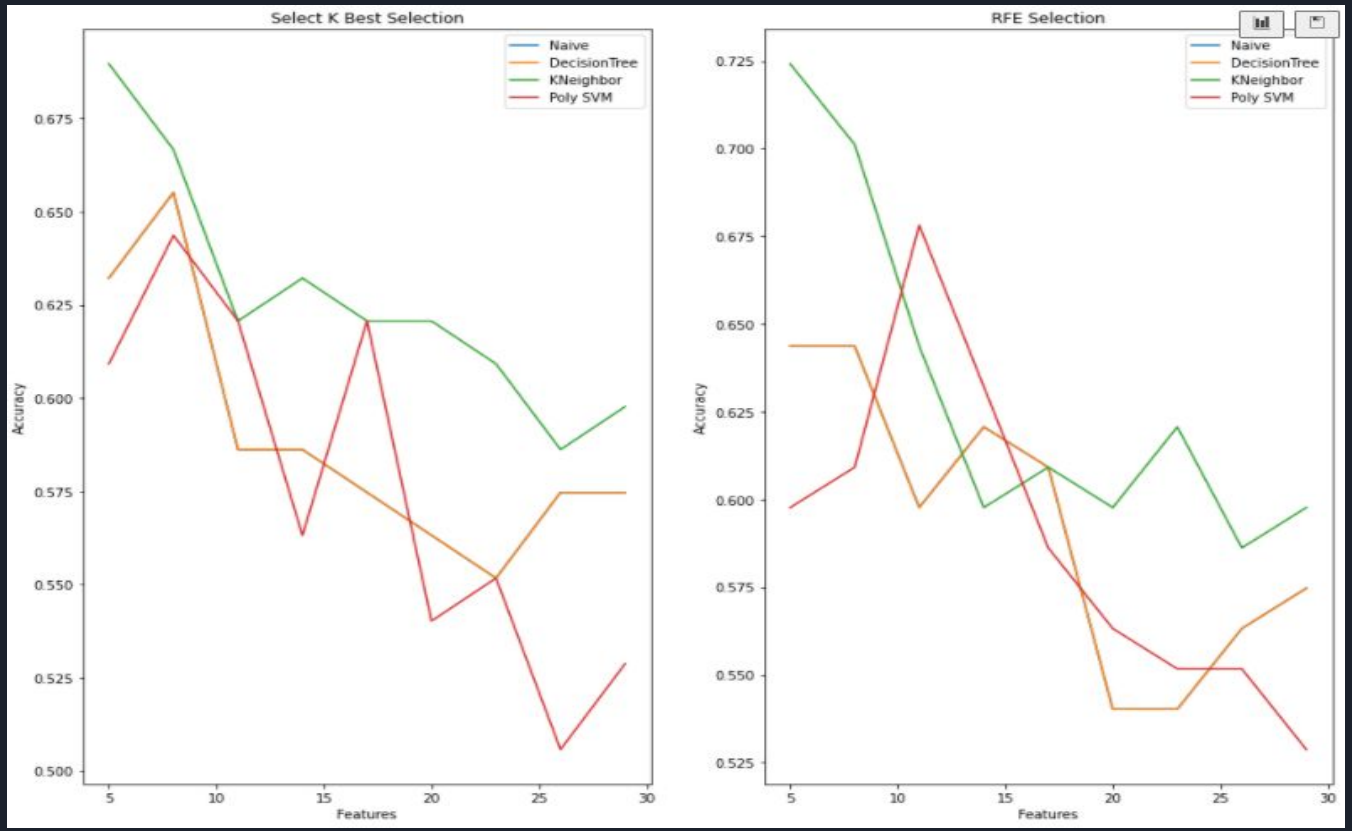
- 32 Attributes
- Grade (1 - 6) compressed to pass and fail
- Course ID dropped

UCI Machine Learning Repository(https://archive.ics.uci.edu/ml/datasets/Higher+Education+Students+Performance+Evaluation+Dataset)

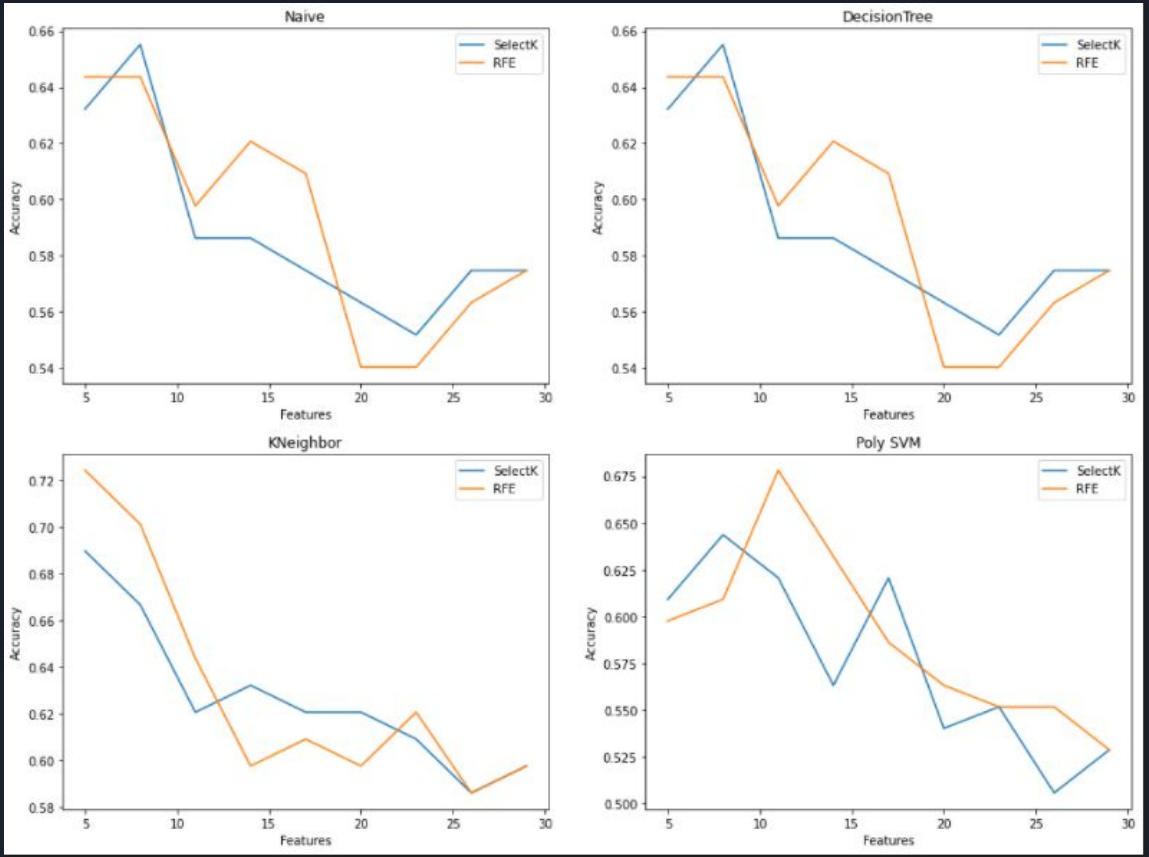
Python Code Tour

Findings

Select K vs RFE Selection Across models



Select K vs RFE per model



Best for predicting performance

Highest Accuracy:

RFE Selection (5 Features):

- Student Age
- Sex
- Salary (Third most)
- Impact of Projects (Second most important)
- GPA Last Semester (Greatest impact)

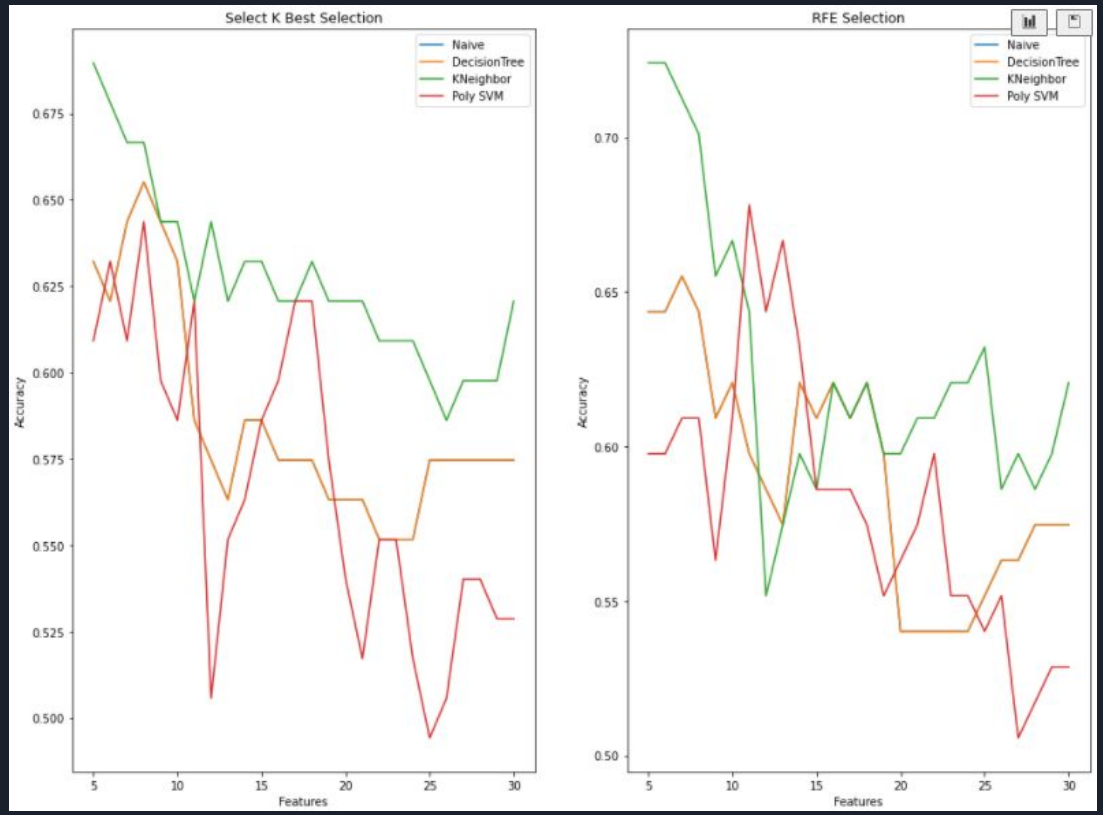
K Neighbor Classification with 5 neighbors

Credits

UCI Machine Learning Repository(https://archive.ics.uci.edu/ml/datasets/Higher+Education+Students+Performance+Evaluation+Dataset)

Thank You

Full Findings



Full Findings (2)