Data set shape is (3673,25)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| MODEL | TRAIN\_R2 | TEST\_R2 | TRAIN\_MSE | TEST\_MSE |
| LINEAR | 0.75 | 0.74 | 0.0077 | 0.0078 |
| POLYNOMIAL | 0.83 | 0.71 | 0.005 | 0.008 |
| POLYNOMIAL WITH LASSO | 0.78 | 0.77 | 0.006 | 0.007 |
| DECISION TREE | 0.87 | 0.63 | 0.003 | 0.011 |
| DECISION TREE TUNED | 0.83 | 0.81 | 0.005 | 0.005 |
| BAGGED LINEAR | 0.75 | 0.74 | 0.008 | 0.007 |
| BAGGED POLYNOMIAL | 0.74 | 0.72 | 0.008 | 0.008 |
| RANDOM FOREST | 0.81 | 0.76 | 0.005 | 0.007 |
| XGBOOST | 0.99 | 0.78 | 0.006 | 0.0002 |
| XGBOOST TUNED |  |  |  |  |

Features used.

features = ["FTFTPCTPELL", "SAT\_AVG\_ALL","TUITIONFEE1\_OUT","DEBT\_MDN", "UG25ABV", "AVGFACSAL","UGDS","COST4", "ADM\_RATE\_ALL", "UGDS\_ASIAN]

Here are the feature along with the feature important scores.

**FTFTPCTPELL (0.489644)**: Description: Represents the percentage of full-time, first-time students awarded Pell Grants. Interpretation: This is by far the most influential feature, suggesting that the rate at which students receive Pell Grants has a significant effect on completion rates. Pell Grants are awarded based on financial need, so this feature may be indicative of the socioeconomic status of students and its impact on completion rates.

**SAT\_AVG\_ALL (0.134555)**: Description: Average SAT scores of all enrolled students. Interpretation: Higher average SAT scores typically indicate a more academically prepared student body, which can influence completion rates. Institutions with higher SAT averages might have more rigorous admissions standards and attract students more likely to complete their degrees.

**TUITIONFEE1\_OUT (0.050522)**: Description: Out-of-state tuition fee. Interpretation: The cost of attending an institution for out-of-state students might affect completion rates. Higher tuition fees might deter some students from completing their degrees, especially if they face financial hardships.

**DEBT\_MDN (0.032774)**: Description: Median debt of students. Interpretation: The amount of debt students take on can influence their decision to continue or drop out. Higher median debt might indicate financial strains on students, potentially affecting completion rates.

**UG25ABV (0.031669)**: Description: Percentage of undergraduates aged 25 and above. Interpretation: Older students might have different life responsibilities (e.g., work, family) that can impact their ability to complete a degree within the expected time. This feature captures the proportion of such non-traditional students.

**AVGFACSAL (0.023334)**: Description: Average faculty salary. Interpretation: Institutions with higher faculty salaries might attract more qualified and experienced faculty, potentially leading to better academic support and thus influencing completion rates.

**UGDS (0.022687)**: Description: Number of undergraduate students enrolled. Interpretation: The size of the student body might affect the institution's resources, class sizes, and student-to-faculty ratios, which can all influence completion rates.

**COST4 (0.021932)**: Description: Average cost for a 4-year program. Interpretation: Similar to tuition fees, the overall cost of a 4-year program can impact a student's ability to complete the program, especially if they face financial challenges.

**ADM\_RATE\_ALL (0.021528)**: Description: Overall admission rate. Interpretation: Institutions with lower admission rates are more selective. The selectivity might correlate with student preparedness and resources available, influencing completion rates.

**UGDS\_ASIAN (0.021096)**: Description: Percentage of Asian undergraduate students. Interpretation: This represents the diversity of the student body. Different demographics might have varying cultural, economic, or academic backgrounds that can impact completion rates.