

Project 11:

Predicting Purdue's Performance in the NCAA Tournament Bracket

Business and Data Understanding

Project Lead: Professor Petros Drineas

Members: Sarah Firestone, Zachary Hanson,

Kuba Bal, Saul Means, Avi Khurana

Business Understanding

Background

Problem: Predict how Purdue will perform in the NCAA tournament

Impact: Viewership, ticket sales, national ranking, improve Purdue's basketball program

Currently: Rely on coaching, real time analysis and video footage

Our Solution: Create a model, predict number of games Purdue wins

Business Understanding

Project Plan

Aim: 4 month project

- **Month 1:** data sources, data familiarity
- **Month 2-3:** data preparation, method selection
- **Month 4:** model adjustment, result analysis, and documentation

Data Understanding

Describing the Data

- 2 main data sources
 - Kaggle
 - Kenpom
- Combined both datasets
- Target Value
- Data from 2013-2024
- Removed all COVID data



-
- 1) 2013-2024 table (excluding 2020):
 1. Season, Team Name is the primary key of the table
 2. This table is in the CSV format
 3. de_TOPct: defensive turnover percent
 4. de_ORPct: defensive offensive rebound percentage
 5. de_FTRate: defensive free throw rate
 6. o_eFGPct: offensive efficiency field goal percentage
 7. o_TOPct: offensive turnover percentage
 8. o_ORPct: offensive offensive rebound percentage
 9. o_FTRate: offensive free throw rate
 10. tempo: average number of possessions per game
 11. Adjtempo: adjusted tempo
 12. OE: Offensive efficiency
 13. AdjOE: Adjusted offensive efficiency
 14. DE: Defensive efficiency
 15. AdjDE: Defensive efficiency
 16. AdjEM: Adjusted efficiency
 18. Off_2: Percentage of scored possessions where the offense scored 2 points.
 19. Off_3: Percentage of scored possessions where the offense scored 3 points.
 21. Def_2: Percentage of scored possessions where the defense allowed 2 points.
 22. Def_3: Percentage of scored possessions where the defense allowed 3 points.
 23. Conf: Conference
 24. G: Games
 25. W: Wins
 26. 2P_O: Offensive 2 pointer percentage
 27. 2P_D: Defensive 2 pointer percentage
 28. 3P_O: Offensive 3 pointer percentage
 29. 3P_D: Defensive 3 pointer percentage
 30. WAB: Wins above bubble
 - 31: Postseason: round team got to in march madness
 - 31: Seed: Seed in march madness

Data Understanding

- All relevant attributes are numeric values
- No missing values
- Defining each attribute
- Dataset contained over 50 attributes
- Need to drop attributes

