Project 11:

Predicting Purdue's Performance in the NCAA Tournament Bracket

Evaluation

Project Lead: Professor Petros Drineas, Christos Boutsikas

Members: Kuba Bal, Saul Means, Sarah Firestone,

Zachary Hanson, Avi Khurana

PURDUE UNIVERSITY

Business Understanding

Recap

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

Success: If within an error of 1 round

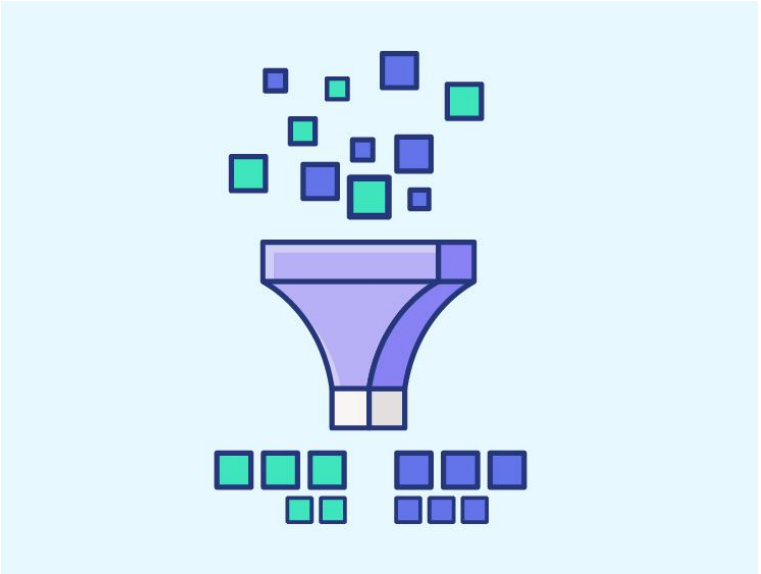
Photo of Shocking 1st Round Exit in 2023

PURDUE UNIVERSITY

Data Preparation

Updates and Adjustments

- Manual adjustments to 2024 data
- Addition of Las Vegas odds
- Only modeling on 2021 and onwards
- Only using 9 columns to train
    a. Kaggle vs. KenPom



PURDUE UNIVERSITY

Modeling Recap

Background

4 Models: Classification

1) Support Vector Machine

2) K Nearest Neighbors

3) Multinomial Logistic Regression

4) Random Forest

PURDUE UNIVERSITY

Recap: Model Predictions

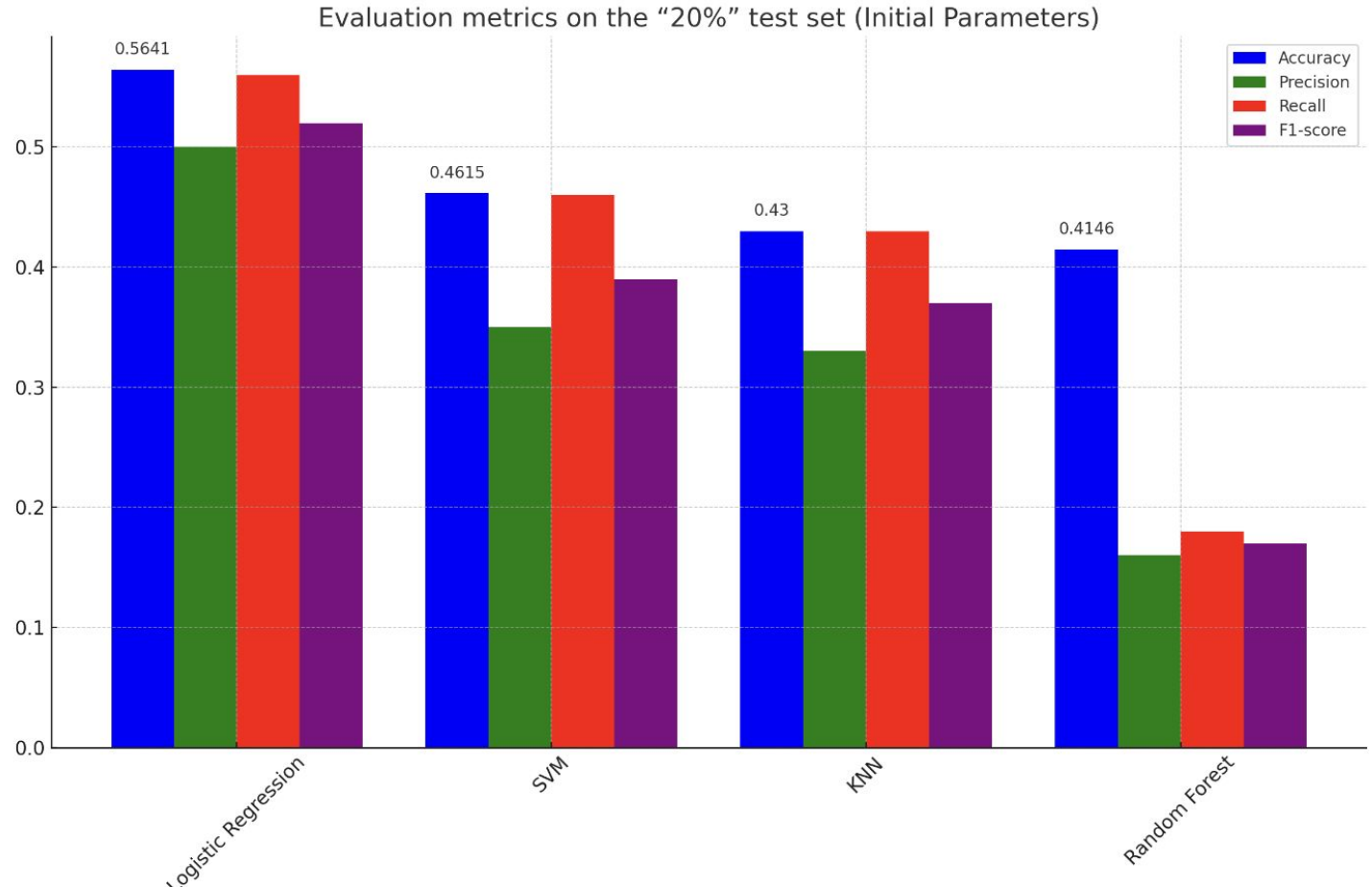
Logistic Regression, SVM, KNN

Random Forest

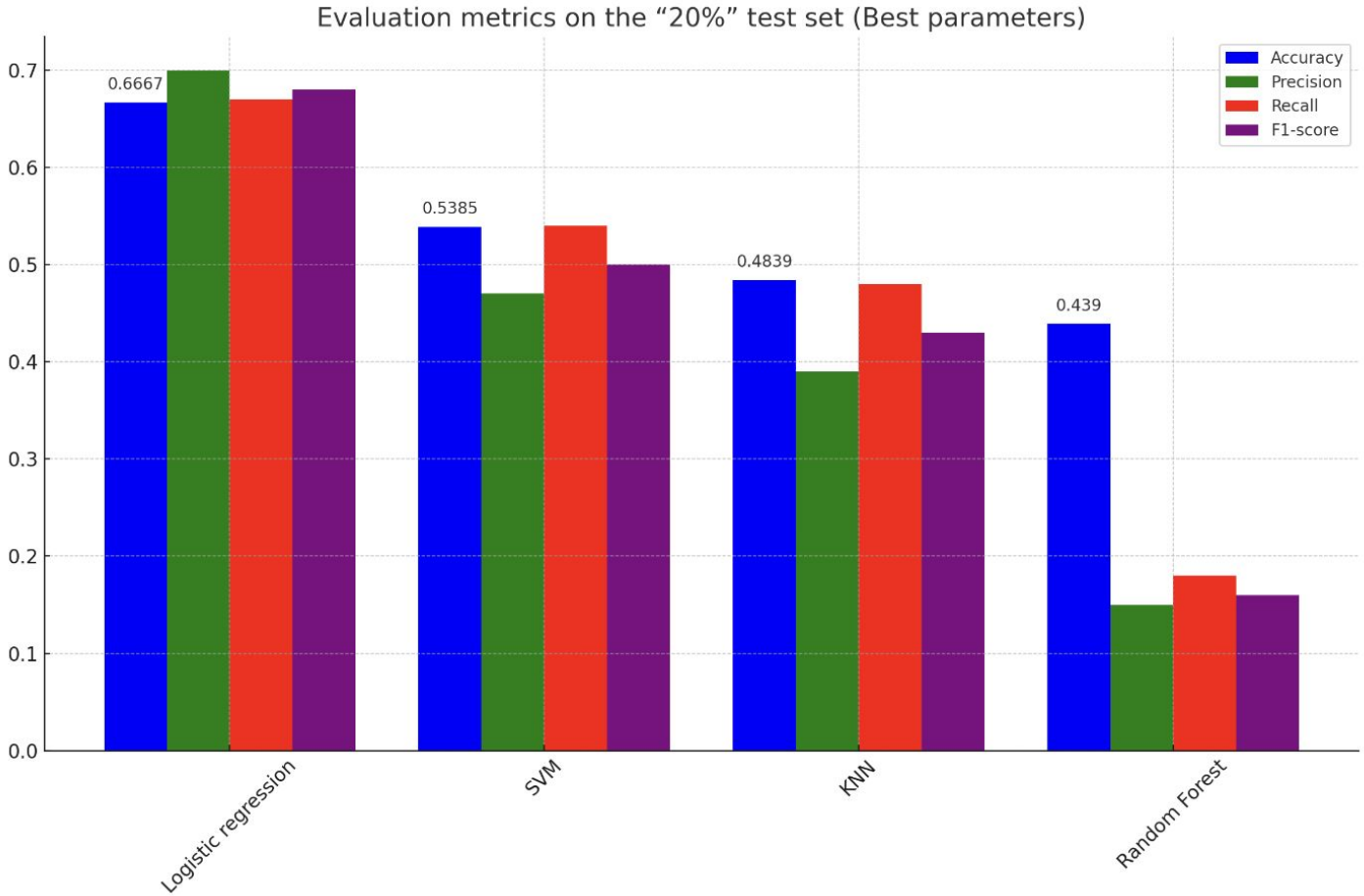
2nd place

PURDUE UNIVERSITY

Modeling Methods Outcomes - *Initial Parameters*



Modeling Methods Outcomes - *Best Parameters*



Evaluation Takeaways

Complexity of the Tournament

- Known for numerous upsets
- “Cinderella Teams” defy statistical models and expert opinions
- “... three sets of [flipping a coin 63 times] and my average scores were better than the average of study participants.” (University of Michigan)
  - https://record.umich.edu/articles/coin-toss-best-bet-march-madness-bracket-success/

PURDUE UNIVERSITY

Evaluation Takeaways

- *"... the combination of modest statistical methods with informative data can meet or exceed the accuracy of more complex models when it comes to predicting the NCAA men's basketball tournament."* (Journal of Quantitative Analysis in Sports)
  - https://doi.org/10.48550/arXiv.1412.0248

PURDUE UNIVERSITY

Next Steps

Predict outcomes of NCAA tournament game-by-game

Data Preparation: each row in the dataset represents a single game in the tournament

- Target Variable: 'OUTCOME' - 0 if Team_1 wins, 1 if Team_2 wins

Season	Team_1	Team_2	Tempo_1	Tempo_2	SEED_1	SEED_2	...	ROUND	OUTCOME
2023	Purdue	Fairleigh Dick	64.2011	69.2289	1	16	...	R64	1

Modeling: Logistic Regression, Random Forest, SVM

PURDUE UNIVERSITY

Thank you!

PURDUE UNIVERSITY®