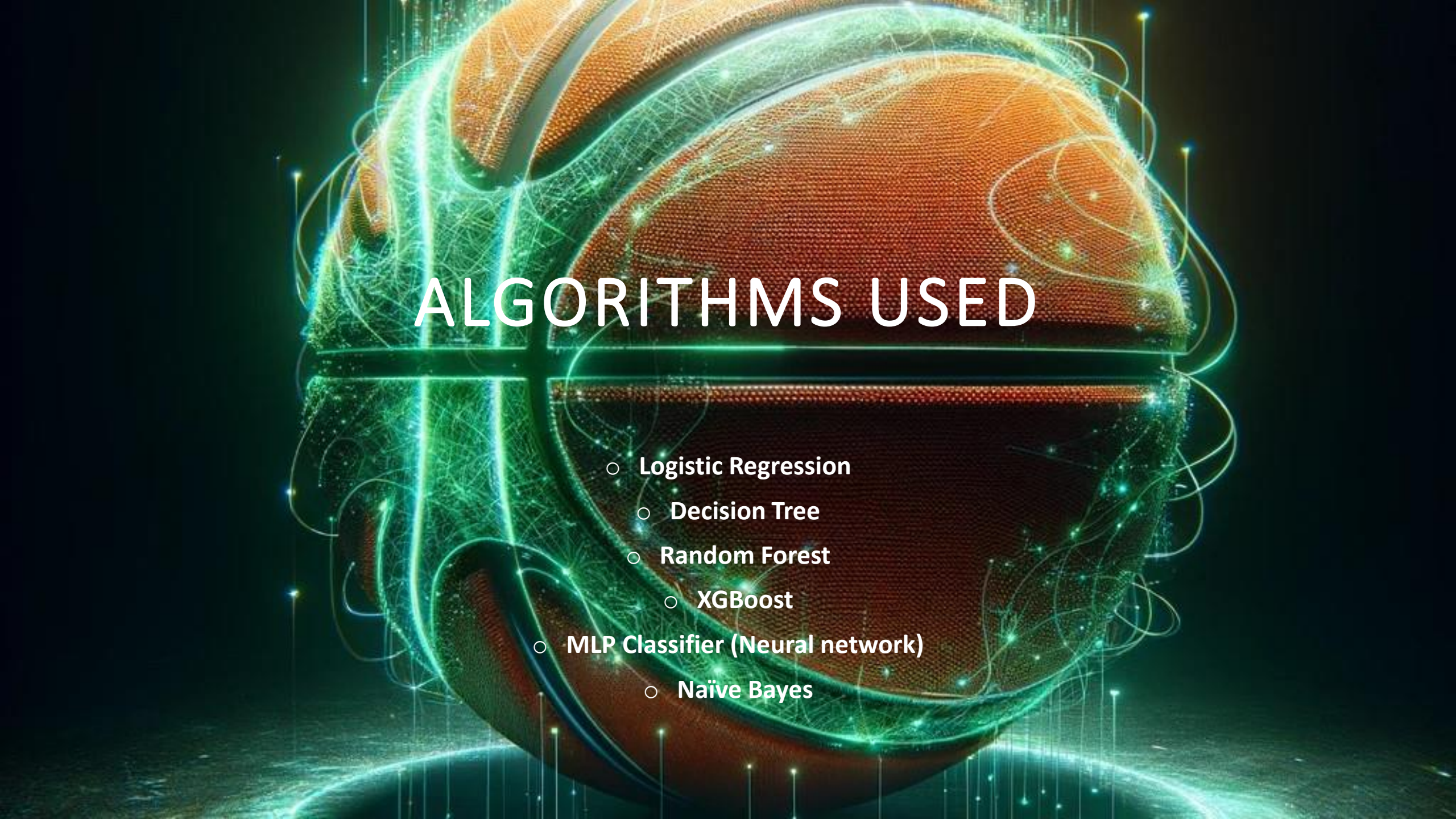
MARCH MACHINE LEARNING MANIA 2024

FORECASTING THE 2024 MEN'S COLLEGE BASKETBALL TOURNAMENTS

THE PROBLEM

OUR GOAL IS TO PREDICT THE RESULTS OF THE MEN'S 2024 COLLEGE BASKETBALL TOURNAMENTS BY SUBMITTING A COLLECTION OF BRACKETS INFORMED BY HISTORICAL DATA, WHICH WILL BE A CLASSIFICATION PROBLEM.



ALGORITHMS USED

- Logistic Regression
- Decision Tree
- Random Forest
- XGBoost
- MLP Classifier (Neural network)
- Naïve Bayes

SIZE ON THE DISK:

712MB

NUMBER OF TRAINING INSTANCES:

20

NUMBER OF FEATURES:

60-180

RANGE:

2

INFORMATION ON THE DATASET

HOW THE ALGORITHMS PERFORMED

THE BASELINE

THE INITIAL AVERAGE PERFORMANCE – LOGISTIC REGRESSION

- The initial assessment of the average performance of the Logistic Regression algorithm on our dataset was 0.73.
- The initial feature configuration was a test using the model as a general test.

ALGORITHM TRAINING & CONTINUOUS EVALUATION

ADDITIONAL ALGORITHMS USED

ALGORITHMS USED - CLASSIFICATION

Top Performers

- Logistic Regression – 77.3%
- MLP Classifier – 77.2%
- XGBoost – 76.0%

Worst Performers

- Naïve Bayes – 75.3%
- Random Forest – 75.1%
- Decision Tree – 68.0%

INTERPRETING THE TOP PERFORMER

- OUR TOP PERFORMING ALGORITHM WAS LOGISTIC REGRESSION, WITH AN ACCURACY OF 77.3%.
- TOP PERFORMING CONFIGURATIONS
  - LOGISTIC REGRESSION WITHOUT RATINGS
  - MLP CLASSIFICATION WITH RATINGS
  - XGBOOST WITH RATINGS

AVERAGES FOR EACH ALGORITHM

Data Variants	Averages	Averages with Ratings	Average of Past 10 Games	Average of Past 10 Games W Rating
Naïve Bayes	0.694963	0.717608	0.743741	0.752905
Logistical Regression	0.735479	0.730485	0.776097	0.772723
Decision Tree	0.642981	0.644373	0.692437	0.680367
Random Forest	0.726145	0.713783	0.756219	0.751237
XGBoost	0.719841	0.718567	0.755495	0.760408
MLP	0.731234	0.729159	0.770152	0.772158


TRICKS

THANK YOU

Project 1

Team: The Overfitting Overlords Taking this to Vegas.

Members: Anton Maynard, Larry Jones, & Kenneth Mitchell

Presentation Date: April 4th, 2024