NBA CAREER LONGEVITY

PREDICTIVE ANALYSIS USING MACHINE LEARNING

OBJECTIVE

CLEANING THE DATA

PROBLEMS ENCOUNTERED:

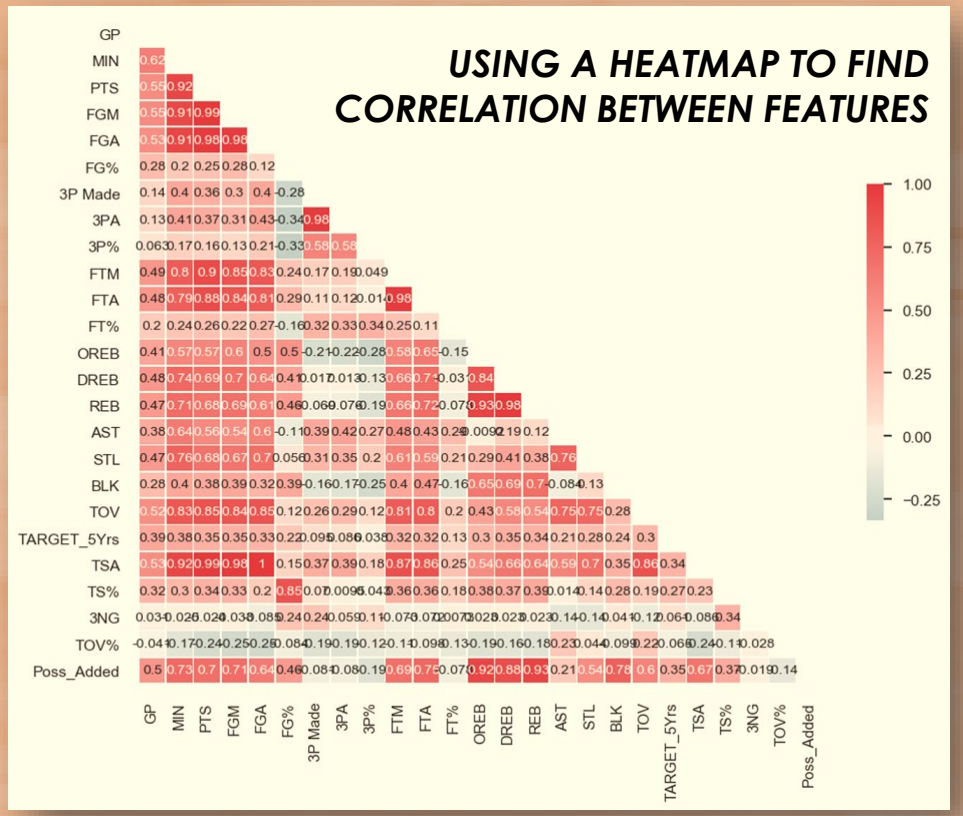
- DUPLICATION
- MISCLASSIFICATIONS
- INCOMPLETE & MISSING DATA
- RECENT DRAFTEES INTRODUCING "NOISE"

EXPLORING THE DATA

ESTABLISHING THE "BASELINE"

CREATING NEW FEATURES

CONDUCTING A "HEAT CHECK"

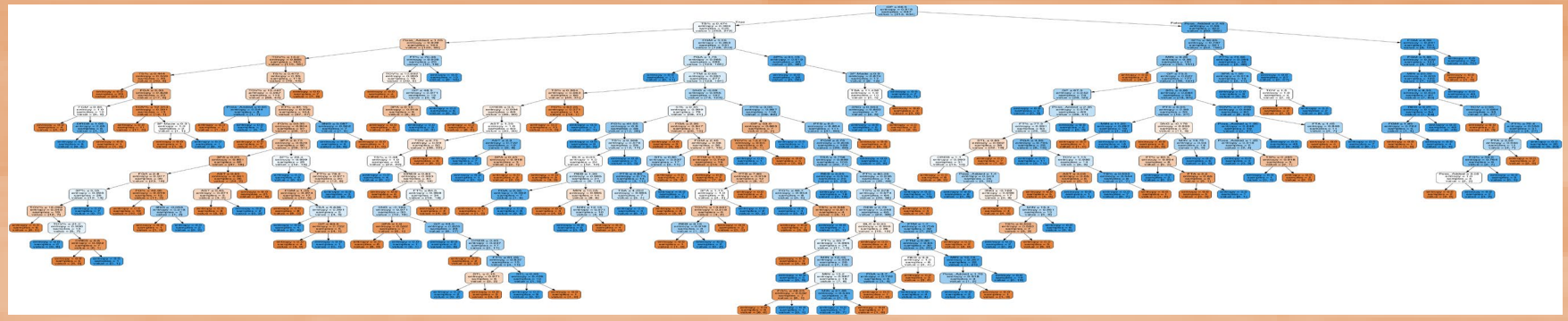


MODEL COMPARISON

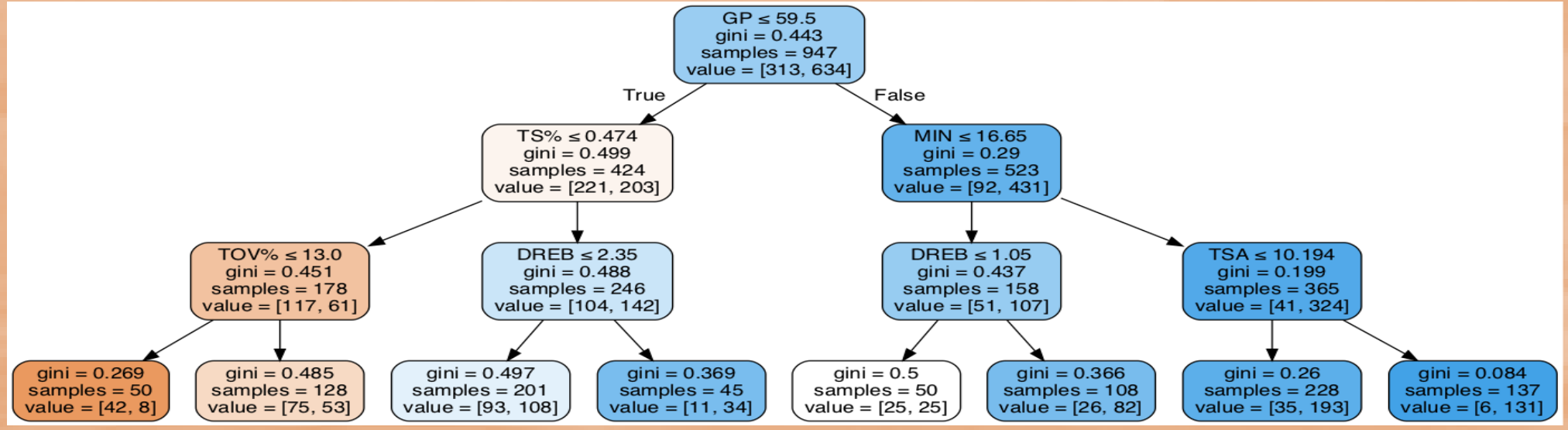
FITTING VARIOUS CLASSIFICATION MODELS WITH DEFAULT SETTINGS

MODEL NAME	ACCURACY SCORE	F1 SCORE	COMPUTATIONAL TIME
Logistic Regression	0.7342	0.8117	0.0214
XGBoost	0.7342	0.8082	0.1222
Random Forest	0.7247	0.7953	0.0280
Gradient Boosting	0.7215	0.8044	0.1787
K-Nearest Neighbors	0.6930	0.7749	0.0091
AdaBoost	0.6899	0.7752	0.2113
Decision Tree	0.6677	0.7518	0.0131

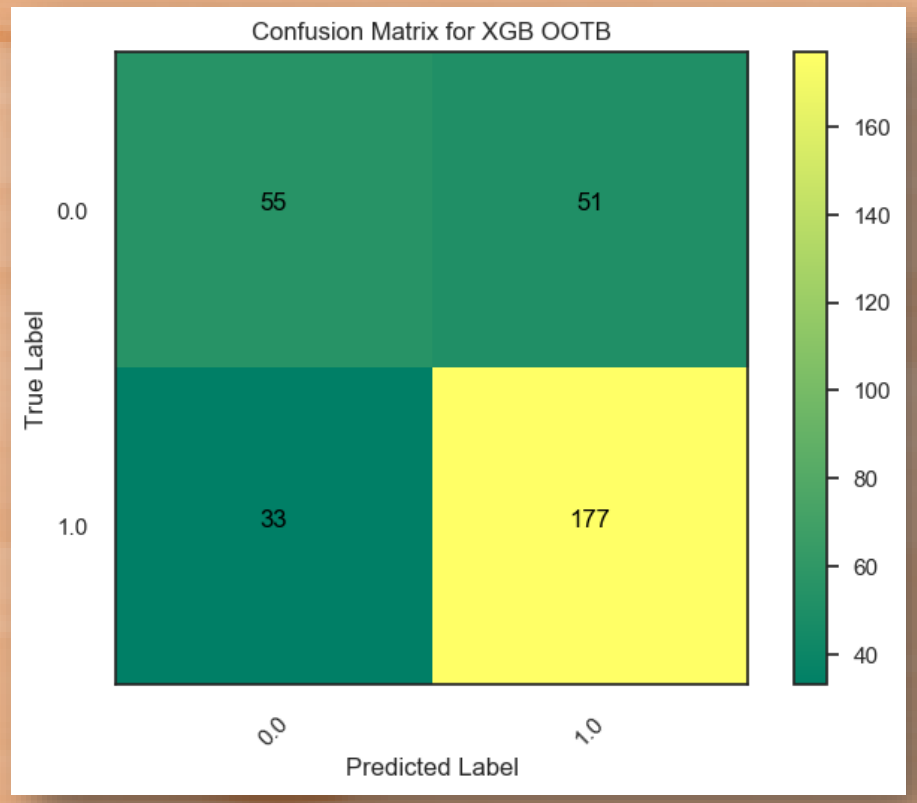
A "SIMPLE" MODEL: DECISION TREE

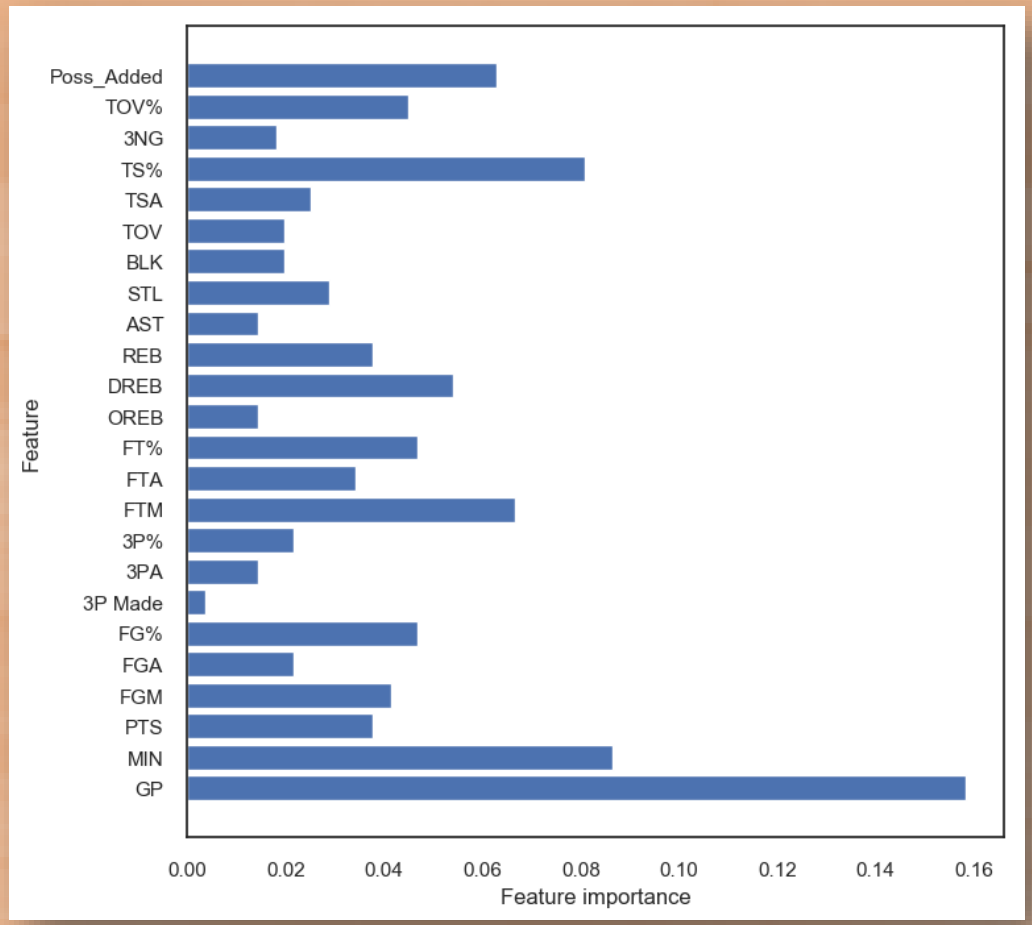


"TUNING" THE MODEL TO FIND THE BEST FIT



EVALUATING A COMPLEX "ENSEMBLE" MODEL: XGBOOST





INTERPRETING THE FINAL MODEL

- WE CAN CORRECTLY CLASSIFY ABOUT 75% OF THE DATA THAT WAS SET ASIDE FOR TESTING PURPOSES
  - 8% IMPROVEMENT OVER THE BASELINE
- FEATURE IMPORTANCE HELPS US DETERMINE WHICH FEATURES/PLAYER ATTRIBUTES ARE MOST PROMINENTLY CONSIDERED
- CAN BE UTILIZED BY NBA FRONT OFFICES, AGENTS, PLAYERS, AND FANS ALIKE