



NBA Standalone Game Performance, Player Positional Classification



Can an NBA player's standalone game statline accurately predict their position?

Alex Salce | MATH574M Final Project December 5th, 2023

Problem Statement



Given an NBA player's single game performance statline, how accurately can ML model predict what position they play?

PG	DATE	TM	OPP	SCORE	MIN	FGM	FGA	FG%	3PM	3PA	3P%	FTM	FTA	FT%
	Sun 11/26	PHX	@ NYK	W 116-113	38	10	25	40.0	1	4	25.0	7	8	87.5



Original Motivation

- I enjoy Basketball!
- Are certain positions more distinct by in-game performance?
- What features “fit the mold” for a position?
- Misclassified players, “positionless basketball”



DATA Summary

	FG_PCT	FG3_PCT	FT_PCT	estimatedOffensiveRating	offensiveRating	estimatedDefensiveRating	defensiveRating	estimatedNetRating	netRating	assistPercentage	assistToTurnover	assistRatio	offensiveReboundPercentage	defensiveReboundPercentage	reboundPercentage	turnoverRatio	effectiveFieldGoalPercentage	trueShootingPercentage	usagePercentage
	0.545	0.400	0.000	111.4	116.9	100.1	98.4	11.3	18.6	0.050	0.5	7.1	0.033	0.387	0.213	14.3	0.636	0.636	0.1
	0.600	0.333	1.000	108.3	108.5	131.8	133.3	-23.5	-24.8	0.000	0.0	0.0	0.080	0.250	0.156	0.0	0.650	0.707	0.2
	0.636	0.500	1.000	88.4	86.8	91.0	97.4	-2.6	-10.5	0.000	0.0	0.0	0.154	0.227	0.188	8.3	0.682	0.699	0.2
	0.500	0.500	1.000	145.3	142.9	96.6	103.3	48.7	39.5	0.111	0.0	10.0	0.083	0.200	0.148	0.0	0.625	0.676	0.2
	0.571	0.000	0.000	111.2	124.0	100.0	104.2	11.2	19.8	0.000	0.0	0.0	0.091	0.154	0.125	12.5	0.571	0.571	0.2
	0.400	0.000	0.000	147.5	150.0	170.5	175.0	-22.9	-25.0	0.000	0.0	0.0	0.200	0.000	0.111	0.0	0.400	0.400	0.2
	0.500	0.400	0.000	110.3	111.1	102.9	104.3	7.5	6.9	0.125	1.0	14.3	0.074	0.280	0.173	14.3	0.600	0.600	0.2
	0.200	0.000	0.000	111.3	111.3	100.0	101.3	11.3	18.8	0.000	0.0	0.0	0.000	0.000	0.000	0.0	0.000	0.000	0.1
	0.000	0.000	0.000	120.0	120.0	102.0	102.0	-55.0	-52.0	0.000	0.0	0.0	0.500	0.000	0.000	0.0	0.000	0.000	0.321
	0.200	0.000	0.000	111.3	111.3	100.0	101.3	11.3	18.8	0.000	0.0	0.0	0.000	0.000	0.000	0.0	0.000	0.000	0.318

Source

All stats queried and assembled from nba_api using Python, output CSV (~85MB)



[nba api \[GitHub\]](#)
[nba api/stats/endpoints/](#)

- [PlayerGameLog](#)
- [BoxScoreHustleV2](#)
- [BoxScoreAdvancedV3](#)
- [BoxScorePlayerTrackV3](#)



Cleanup

- Redundant features
- Scale all non-percentage stats by time played in game
- Remove any physical characteristic features

Features ["p" dimension]

- p=76 features
- 19 standard boxscore stats
 - 22 “advanced” stats
 - 15 “hustle” stats
 - 20 “player tracking” stats

Observations ["n" dimension]

- n=163,230 player statlines
 - Individual NBA game performances 2006-2023
 - Regular Season games only

Other notes

- Some stat pairs are correlated by nature, certain redundancies exist (future cleanup)

Season	n
2023-24	3,255
2022-23	25,774
2021-22	22,704
2020-21	18,125
2019-20	15,836
2018-19	15,507
2017-18	12,578

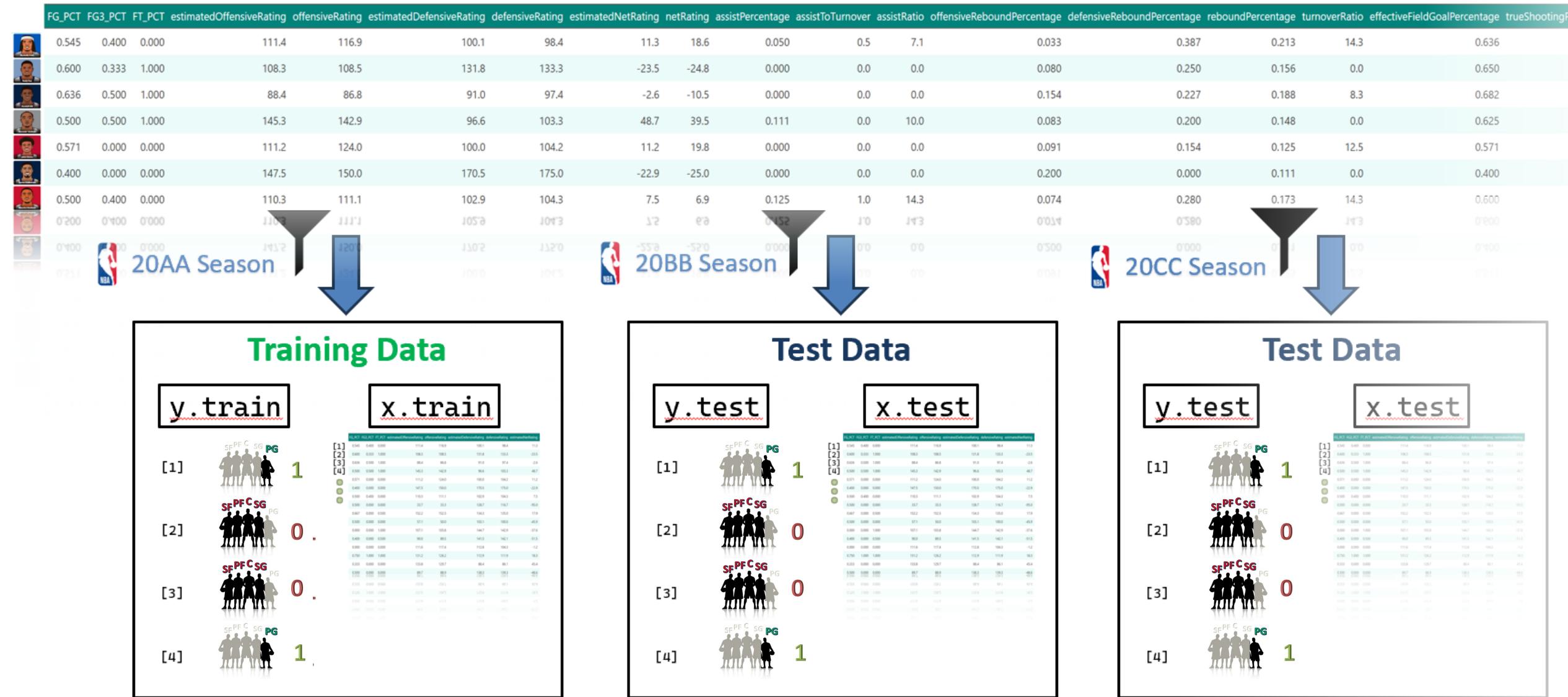


THE UNIVERSITY
OF ARIZONA

Training and Test Data

Generate Training Data

- Select Training Season
 $\{2017, 2018, 2019, 2020, 2021, \textcolor{red}{2022}, 2023\}$
- Filter, create `x.train df`
- Select position classification
 - [1] Point Guard
 - [2] Shooting guard
 - [3] Small Forward
 - [4] Power Forward
 - [5] Center
- Reassign class data to 1/0 for position selection
- Create `y.train df`



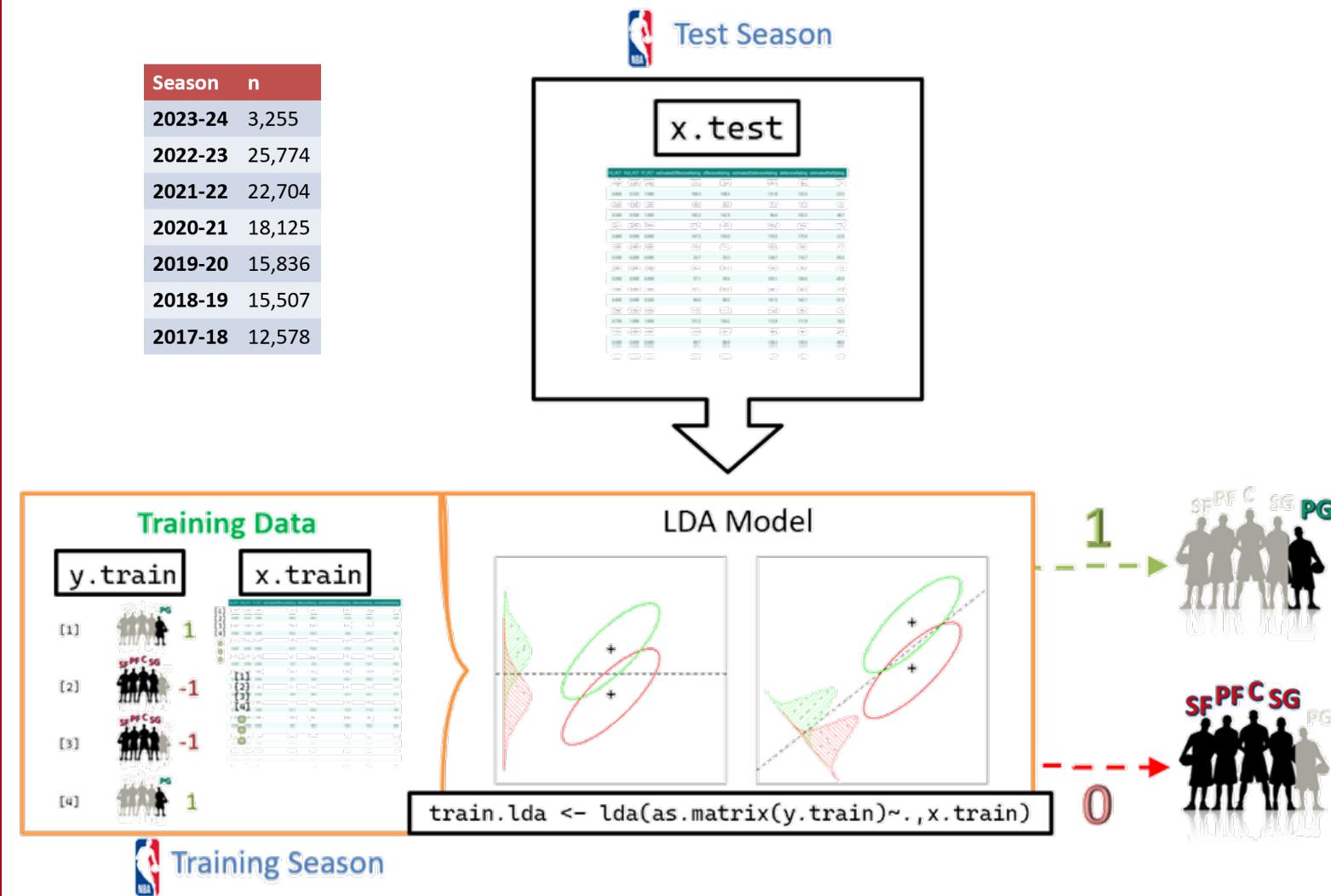
Generate Test Data

- Select Training Season
 $\{2017, 2018, 2019, 2020, 2021, \textcolor{red}{2022}, 2023\}$
- Create `x.train df`
- Reassign class data to 1/-1 from position selection
- Create `y.train df`

- **Binary Classification** used (rather than multiclass)
- **Train by individual season** to compare to other Test seasons individually

LDA Model Results

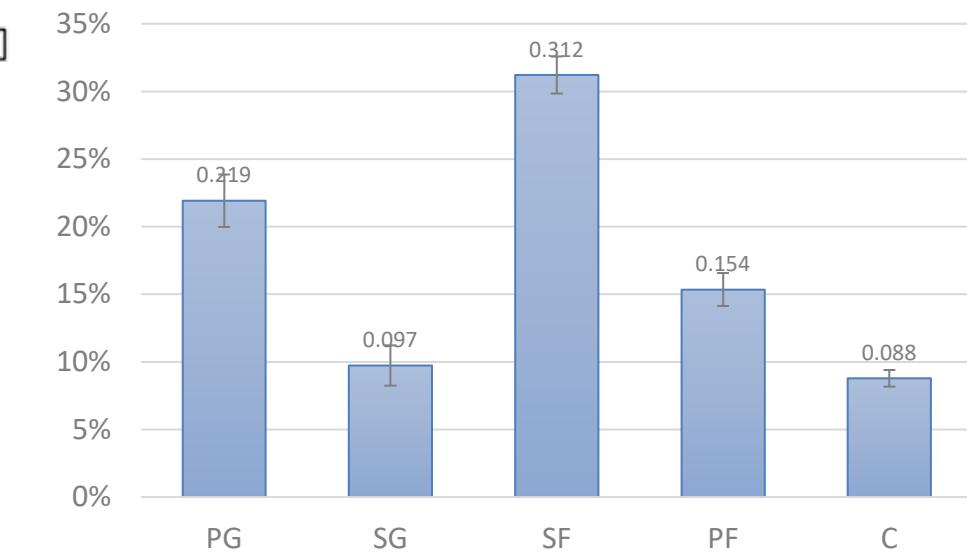
Season	n
2023-24	3,255
2022-23	25,774
2021-22	22,704
2020-21	18,125
2019-20	15,836
2018-19	15,507
2017-18	12,578



Example data: Point Guard

PG	X5CVErr	trainErr	X2023	X2022	X2021	X2020	X2019	X2018	X2017
2023	22.215%	21.997%	NA	23.164%	23.687%	23.349%	23.810%	25.438%	27.035%
2022	20.675%	22.309%	22.674%	NA	22.387%	22.647%	22.755%	23.803%	24.692%
2021	21.595%	21.402%	22.692%	21.846%	NA	21.948%	21.921%	22.705%	23.586%
2020	21.115%	21.175%	22.792%	21.782%	21.550%	NA	21.727%	22.389%	23.492%
2019	20.885%	21.325%	23.061%	22.228%	21.798%	22.038%	NA	21.817%	22.594%
2018	18.968%	18.953%	21.487%	20.359%	19.901%	19.926%	19.546%	NA	19.823%
2017	16.993%	17.244%	20.003%	18.699%	18.413%	19.375%	18.493%	17.332%	NA
Summary					avg trainErr				
PG	21.923%	1.939%	20.629%	1.705%					

Test Error Summary

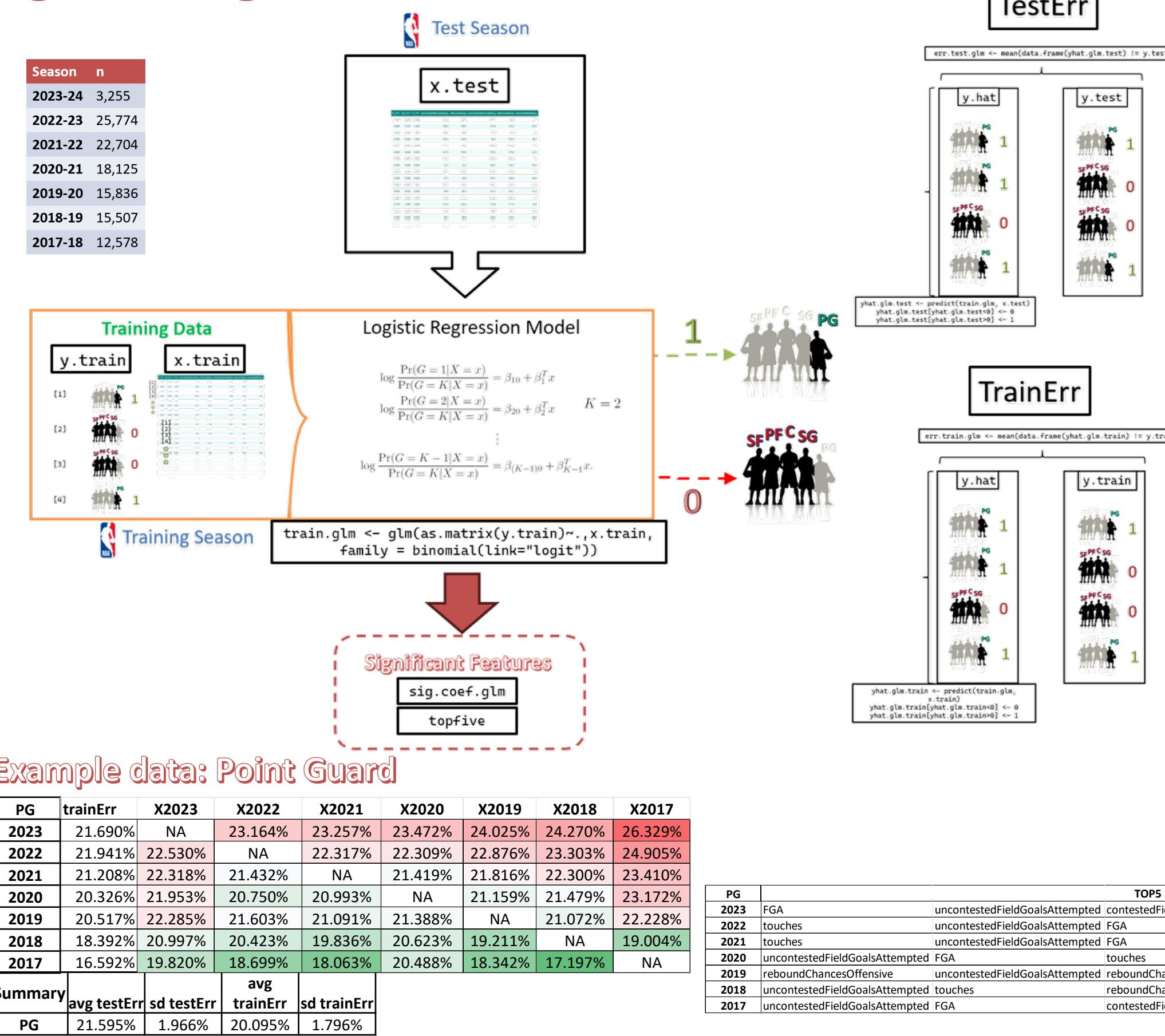


- Best prediction for SGs and Cs
- Worst prediction for SFs

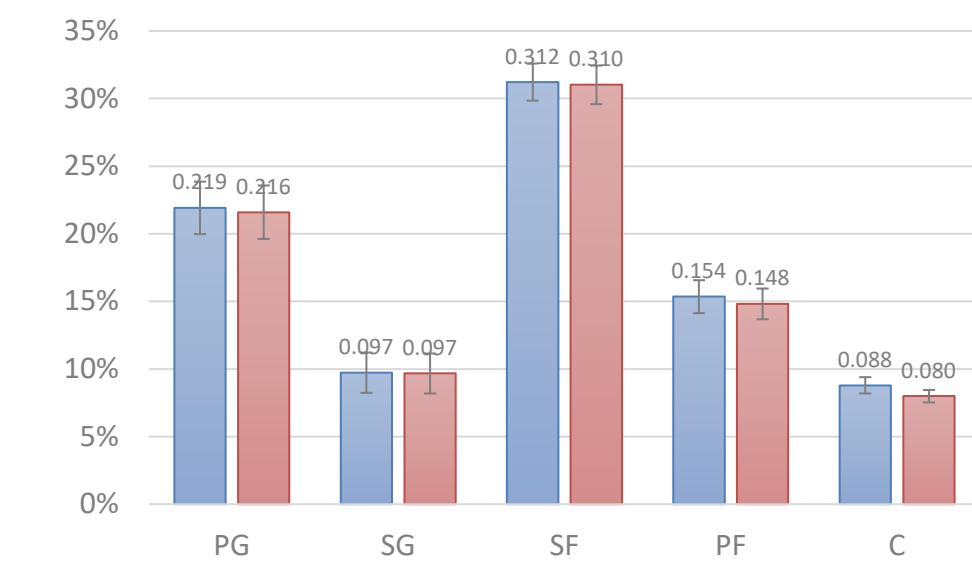
Train Error Summary



Logistic Regression Model Results

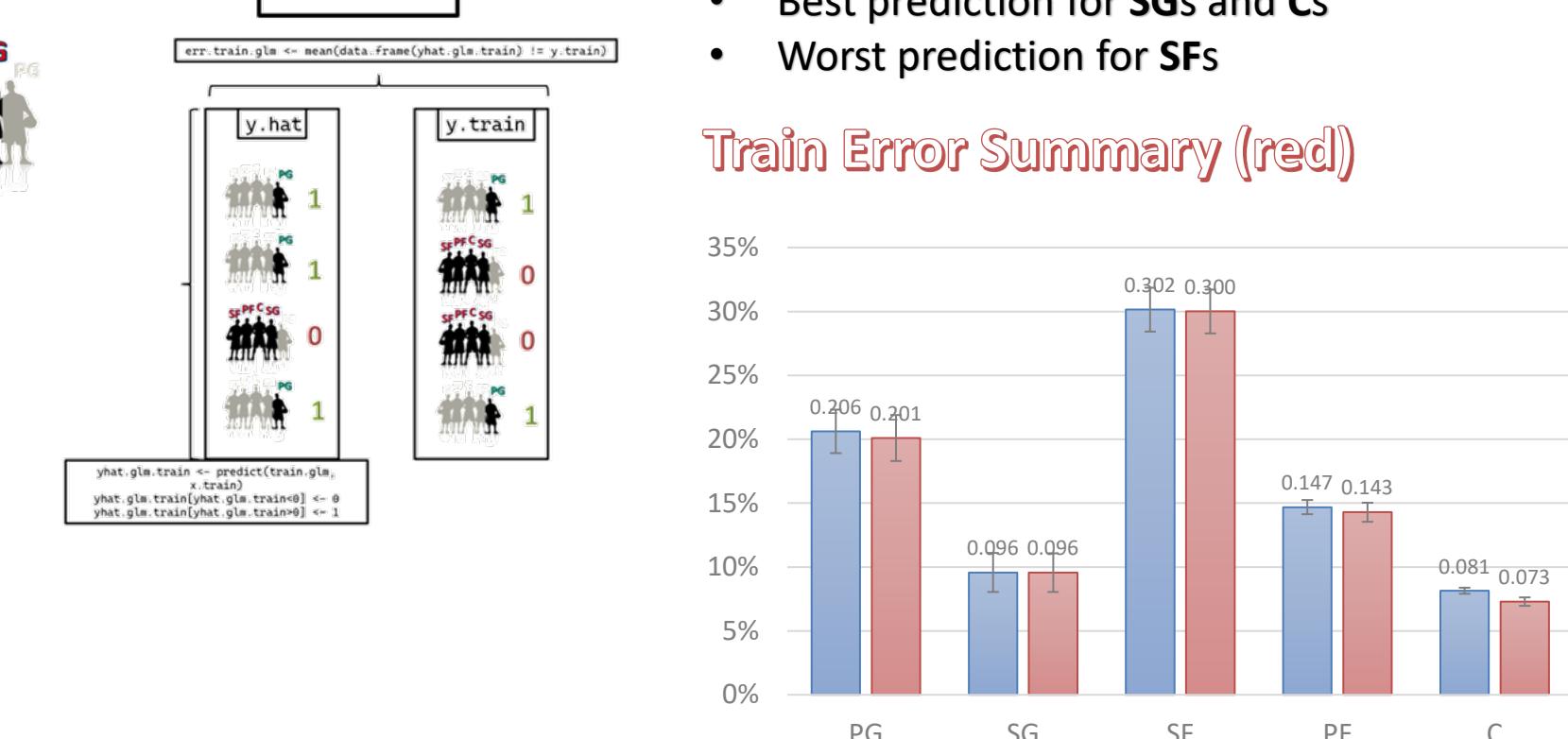


Test Error Summary (red)



- Generally outperforms LDA
- Best prediction for SGs and Cs
- Worst prediction for SFs

Train Error Summary (red)

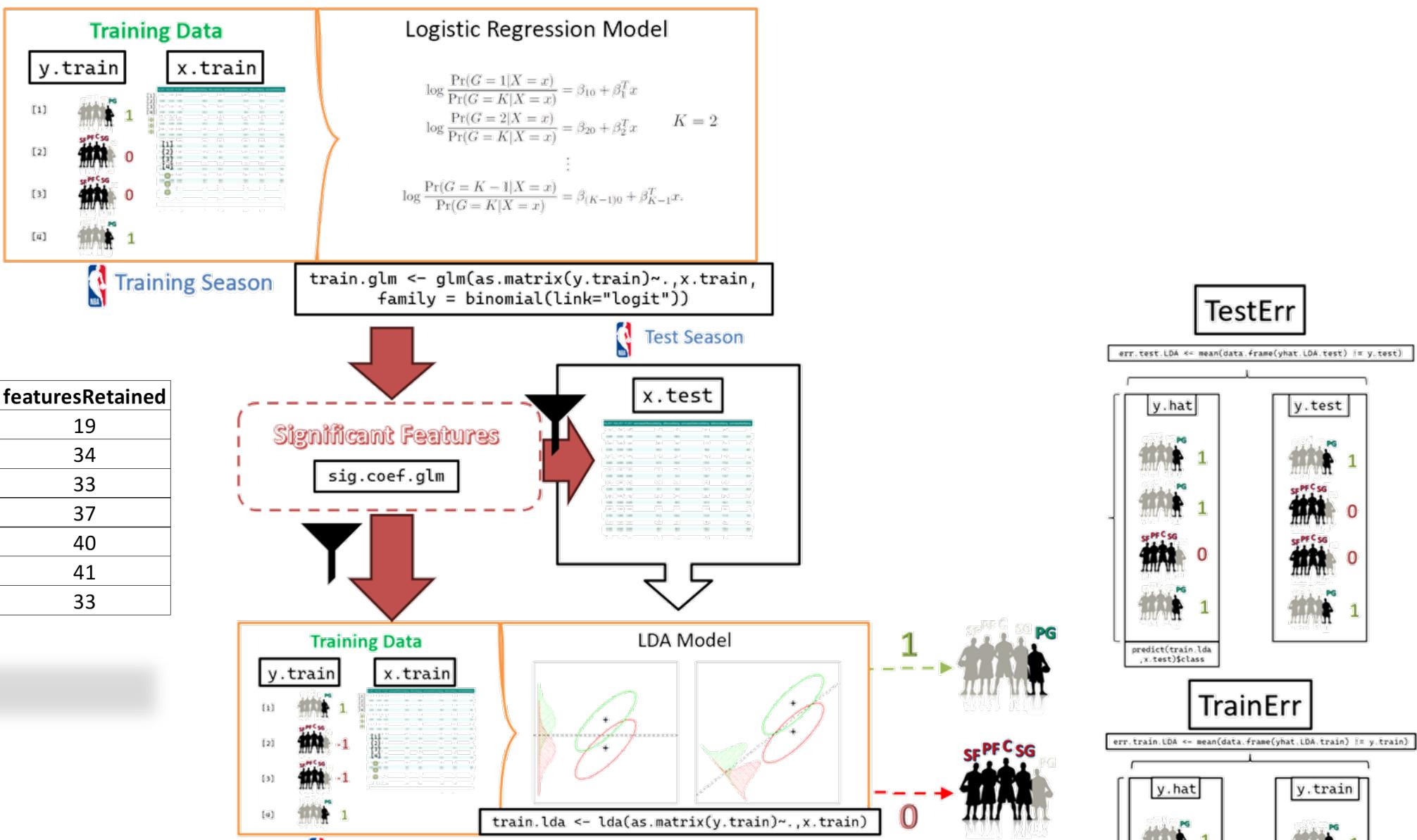


PG	TOP5 Features				
2023	FGA	uncontestedFieldGoalsAttempted	contestedFieldGoalsAttempted	touches	deflections
2022	touches	uncontestedFieldGoalsAttempted	FGA	contestedFieldGoalsAttempted	contestedShots2pt
2021	touches	uncontestedFieldGoalsAttempted	FGA	contestedFieldGoalsAttempted	defensiveReboundPercentage
2020	uncontestedFieldGoalsAttempted	FGA	touches	contestedFieldGoalsAttempted	reboundChancesOffensive
2019	reboundChancesOffensive	uncontestedFieldGoalsAttempted	reboundChancesDefensive	reboundChancesTotal	contestedFieldGoalsAttempted
2018	uncontestedFieldGoalsAttempted	touches	reboundChancesOffensive	FGA	contestedFieldGoalsAttempted
2017	uncontestedFieldGoalsAttempted	FGA	contestedFieldGoalsAttempted	touches	reboundChancesOffensive



LDA using Logistic Regression Feature Selection

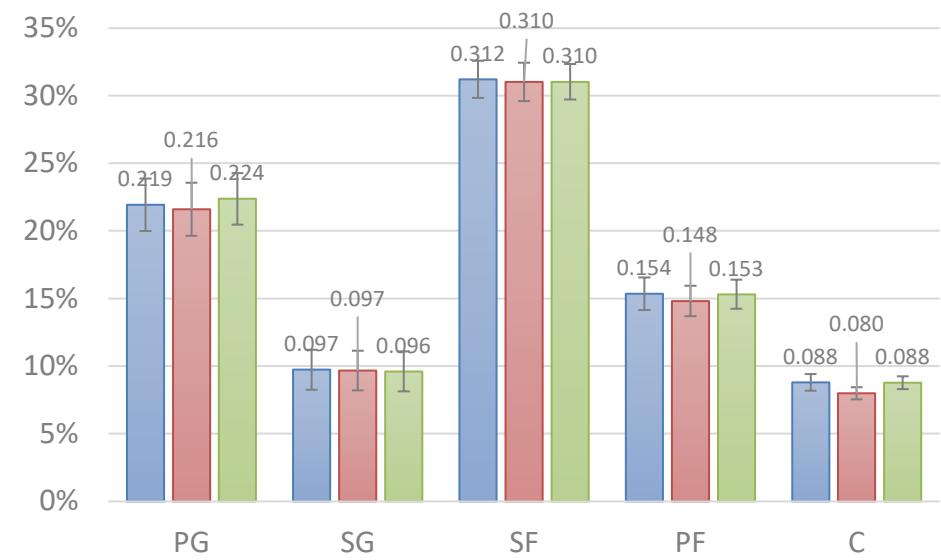
PG	featuresDropped	featuresRetained
2023	52	19
2022	37	34
2021	37	33
2020	34	37
2019	31	40
2018	31	41
2017	38	33



Example data: Point Guard

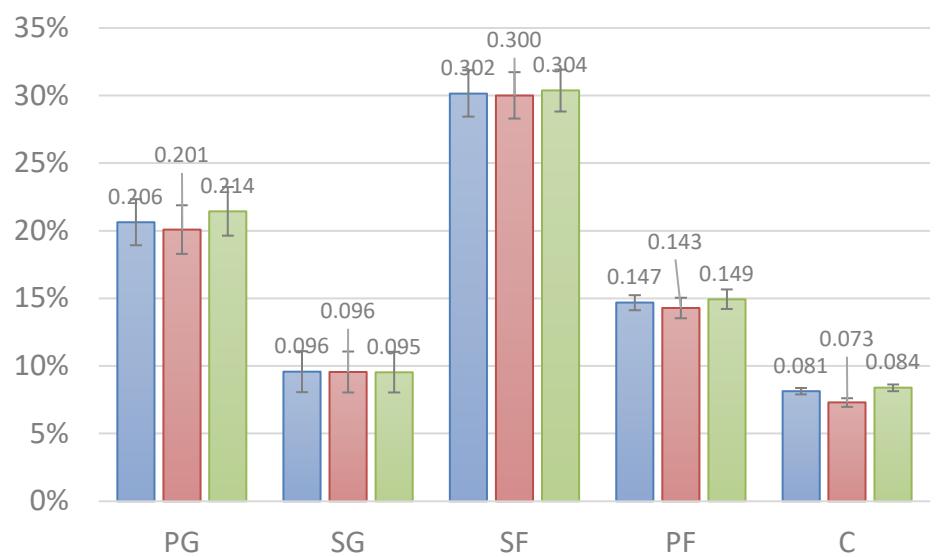
PG	trainErr	X5CVErr	X2023	X2022	X2021	X2020	X2019	X2018	X2017
2023	24.485%	25.108%	NA	23.195%	23.564%	23.472%	23.656%	24.055%	26.390%
2022	22.379%	21.257%	24.827%	NA	22.585%	22.860%	22.697%	23.264%	24.633%
2021	21.344%	21.419%	24.423%	21.939%	NA	21.992%	22.137%	22.190%	23.370%
2020	21.606%	21.656%	24.811%	22.030%	21.732%	NA	21.854%	22.041%	23.294%
2019	22.001%	21.290%	25.631%	22.569%	21.584%	22.588%	NA	22.127%	23.162%
2018	19.823%	20.206%	25.047%	20.745%	19.978%	20.462%	20.191%	NA	20.920%
2017	18.365%	18.870%	23.438%	19.081%	18.421%	19.526%	18.986%	18.206%	NA
Summary	avg testErr	sd testErr	avg trainErr	sd trainErr					
PG	22.373%	1.903%	21.429%	1.796%					

Test Error Summary (green)



- Generally underperforms original LDA
- Best prediction for SGs and Cs
- Worst prediction for SFs

Train Error Summary (green)



Conclusions

- LDA & GLM perform surprisingly well overall
- Identifies Centers & Shooting Guards best
- Feature selection reduces dimensionality significantly with little cost to LDA performance
- Logistic Regression slightly outperforms LDA



Future...

- Misclassified player statlines
- Analysis of features by year
- Nonparametric methods
- Further Data Refinement
- Advanced stats fidelity
- Other cuts of same data framework



TOP5 Features						featuresDropped	featuresRetained
PG							
2023	FGA	uncontestedFieldGoalsAttempted	contestedFieldGoalsAttempted	touches	deflections	52	19
2022	touches	uncontestedFieldGoalsAttempted	FGA	contestedFieldGoalsAttempted	contestedShots2pt	37	34
2021	touches	uncontestedFieldGoalsAttempted	FGA	contestedFieldGoalsAttempted	defensiveReboundPercentage	37	33
2020	uncontestedFieldGoalsAttempted	FGA	touches	contestedFieldGoalsAttempted	reboundChancesOffensive	34	37
2019	reboundChancesOffensive	uncontestedFieldGoalsAttempted	reboundChancesDefensive	reboundChancesTotal	contestedFieldGoalsAttempted	31	40
2018	uncontestedFieldGoalsAttempted	touches	reboundChancesOffensive	FGA	contestedFieldGoalsAttempted	31	41
2017	uncontestedFieldGoalsAttempted	FGA	contestedFieldGoalsAttempted	touches	reboundChancesOffensive	38	33
SG							
2023	reboundChancesDefensive	reboundChancesTotal	PF	assistRatio	seconds	59	11
2022	seconds	reboundChancesTotal	touches	contestedShots	assistRatio	54	17
2021	contestedShots	estimatedPace	FG3M	offensiveBoxOuts	DREB	59	12
2020	screenAssists	contestedFieldGoalsMade	contestedShots	FG3A	distance	52	20
2019	screenAssists	contestedShots	uncontestedFieldGoalsPercentage	screenAssistPoints	assistRatio	52	20
2018	screenAssists	FG3_PCT	seconds	screenAssistPoints	uncontestedFieldGoalsMade	47	25
2017	seconds	FG3_PCT	AST	screenAssists	secondaryAssists	52	19
SF							
2023	possessions	deflections	FG3M	FG3A	contestedShots2pt	58	13
2022	screenAssists	FG3_PCT	screenAssistPoints	contestedShots	contestedShots2pt	40	31
2021	contestedShots	screenAssists	FG3_PCT	screenAssistPoints	contestedShots2pt	33	37
2020	seconds	screenAssists	screenAssistPoints	FG3_PCT	FG3M	43	28
2019	contestedShots2pt	contestedShots	reboundChancesDefensive	reboundChancesTotal	FG3M	44	28
2018	FG3_PCT	contestedShots2pt	FG3M	reboundChancesDefensive	contestedFieldGoalsAttempted	38	35
2017	uncontestedFieldGoalsAttempted	contestedFieldGoalsAttempted	contestedShots2pt	seconds	FGA	42	29
PF							
2023	uncontestedFieldGoalsAttempted	reboundChancesTotal	FG3_PCT	defensiveReboundPercentage	turnoverRatio	51	19
2022	touches	FGA	defensiveReboundPercentage	passes	AST	44	27
2021	contestedShots	BLK	DREB	uncontestedFieldGoalsAttempted	defensiveReboundPercentage	39	31
2020	contestedShots	passes	DREB	touches	seconds	41	30
2019	contestedShots	PF	DREB	secondaryAssists	passes	49	22
2018	contestedShots	seconds	PF	contestedShots2pt	BLK	48	25
2017	contestedShots	assistRatio	passes	boxOutPlayerRebounds	FG3_PCT	43	29
C							
2023	seconds	assistRatio	FG3A	screenAssists	FGM	59	12
2022	screenAssists	FG3A	FGA	distance	FTM	28	44
2021	contestedShots2pt	FGA	FG3A	contestedFieldGoalsAttempted	uncontestedFieldGoalsAttempted	43	28
2020	usagePercentage	FGA	uncontestedFieldGoalsAttempted	estimatedUsagePercentage	distance	33	38
2019	contestedShots2pt	defensiveReboundPercentage	contestedFieldGoalsAttempted	FTM	uncontestedFieldGoalsAttempted	45	26
2018	contestedShots2pt	distance	speed	FG3A	FGA	39	33
2017	contestedShots2pt	FG3A	FGA	distance	contestedFieldGoalsAttempted	43	28

- GREEN indicates LDA outperforms comparison
- RED indicates compare method outperforms LDA

PG	LDA.TestErr - GLM.TestErr							LDA.TestErr - GLM Feature Selection LDA rerun.TestErr						
	X2023	X2022	X2021	X2020	X2019	X2018	X2017	X2023.1	X2022.1	X2021.1	X2020.1	X2019.1	X2018.1	X2017.1
2023	NA	0.000%	0.430%	-0.123%	-0.215%	1.167%	0.707%	NA	-0.031%	0.123%	-0.123%	0.154%	1.382%	0.645%
2022	0.144%	NA	0.070%	0.338%	-0.120%	0.501%	-0.213%	-2.153%	NA	-0.198%	-0.213%	0.058%	0.539%	0.058%
2021	0.374%	0.414%	NA	0.529%	0.106%	0.405%	0.176%	-1.731%	-0.092%	NA	-0.044%	-0.216%	0.515%	0.216%
2020	0.839%	1.032%	0.557%	NA	0.568%	0.910%	0.320%	-2.019%	-0.248%	-0.182%	NA	-0.127%	0.348%	0.199%
2019	0.777%	0.625%	0.707%	0.650%	NA	0.745%	0.366%	-2.570%	-0.341%	0.215%	-0.549%	NA	-0.309%	-0.568%
2018	0.490%	-0.064%	0.064%	-0.696%	0.335%	NA	0.819%	-3.560%	-0.387%	-0.077%	-0.535%	-0.645%	NA	-1.096%
2017	0.183%	0.000%	0.350%	-1.113%	0.151%	0.135%	NA	-3.435%	-0.382%	-0.008%	-0.151%	-0.493%	-0.875%	NA

SG	LDA.TestErr - GLM.TestErr							LDA.TestErr - GLM Feature Selection LDA rerun.TestErr						
	X2023	X2022	X2021	X2020	X2019	X2018	X2017	X2023.1	X2022.1	X2021.1	X2020.1	X2019.1	X2018.1	X2017.1
2023	NA	0.000%	0.031%	0.184%	0.246%	0.369%	0.031%	NA	0.000%	0.246%	0.215%	0.430%	0.553%	0.123%
2022	-0.016%	NA	0.012%	-0.027%	0.186%	0.233%	-0.062%	0.237%	NA	0.140%	0.113%	0.423%	0.361%	0.070%
2021	0.062%	-0.013%	NA	0.013%	0.154%	0.110%	-0.031%	0.189%	-0.009%	NA	0.092%	0.352%	0.308%	0.048%
2020	-0.011%	0.011%	-0.017%	NA	0.099%	-0.011%	-0.050%	0.110%	0.011%	0.072%	NA	0.177%	0.149%	0.088%
2019	0.063%	259408938	-0.013%	259408951	NA	0.044%	-0.013%	0.107%	0.013%	0.025%	0.051%	NA	0.076%	0.025%
2018	-0.013%	0.000%	0.013%	-0.045%	0.026%	NA	-0.045%	-0.019%	0.013%	0.058%	-0.013%	0.090%	NA	0.071%
2017	0.016%	0.000%	0.000%	895690891	0.024%	895690891	NA	0.040%	0.000%	0.040%	0.032%	0.111%	0.064%	NA

SF	LDA.TestErr - GLM.TestErr							LDA.TestErr - GLM Feature Selection LDA rerun.TestErr						
	X2023	X2022	X2021	X2020	X2019	X2018	X2017	X2023.1	X2022.1	X2021.1	X2020.1	X2019.1	X2018.1	X2017.1
2023	NA	0.215%	-0.154%	-0.061%	-0.061%	0.061%	0.092%	NA	0.123%	-0.154%	0.307%	-0.215%	0.000%	0.707%
2022	-0.147%	NA	0.035%	-0.047%	-0.035%	0.039%	0.085%	0.601%	NA	-0.213%	-0.008%	0.109%	0.109%	0.124%
2021	-0.084%	0.018%	NA	-0.004%	-0.004%	-0.048%	0.269%	0.780%	0.216%	NA	-0.022%	-0.176%	-0.044%	0.374%
2020	-0.276%	0.033%	0.022%	NA	-0.055%	-0.248%	0.348%	1.059%	0.320%	0.072%	NA	-0.017%	-0.072%	0.436%
2019	0.063%	0.170%	-0.006%	0.000%	NA	0.158%	0.101%	0.669%	0.013%	0.000%	0.076%	NA	-0.183%	0.259%
2018	0.039%	0.090%	0.161%	0.032%	0.103%	NA	0.310%	0.761%	0.284%	0.677%	0.193%	-0.284%	NA	-0.032%
2017	-0.040%	0.103%	0.167%	0.111%	0.127%	0.644%	NA	0.135%	0.429%	0.358%	0.103%	-0.135%	-0.064%	NA

PF	LDA.TestErr - GLM.TestErr							LDA.TestErr - GLM Feature Selection LDA rerun.TestErr						
	X2023	X2022	X2021	X2020	X2019	X2018	X2017	X2023.1	X2022.1	X2021.1	X2020.1	X2019.1	X2018.1	X2017.1
2023	NA	0.430%	0.307%	0.430%	0.860%	0.399%	0.307%	NA	0.338%	-0.061%	0.246%	-0.031%	-0.492%	0.369%
2022	0.679%	NA	0.590%	0.745%	0.784%	0.318%	0.113%	-0.155%	NA	757895553	0.175%	-0.404%	-0.365%	-0.198%
2021	0.507%	0.700%	NA	0.806%	0.647%	0.344%	0.132%	0.022%	0.573%	NA	0.123%	-0.277%	-0.326%	-0.291%
2020	0.508%	0.375%	0.491%	NA	0.535%	0.199%	-0.006%	0.210%	0.458%	0.276%	NA	-0.353%	-0.320%	-0.072%
2019	0.038%	0.088%	-0.051%	0.410%	NA	0.290%	0.259%	-0.013%	0.177%	-0.120%	0.095%	NA	-0.202%	0.177%
2018	-0.129%	0.232%	0.019%	0.071%	-0.006%	NA	0.032%	-0.090%	0.593%	0.039%	-0.058%	-0.123%	NA	-0.193%
2017	0.246%	0.970%	0.731%	0.882%	0.716%	0.716%	NA	0.135%	1.169%	0.239%	-0.040%	0.533%	-0.127%	NA

C	LDA.TestErr - GLM.TestErr
---	---------------------------

