Injury Analysis and Availability Prediction Model

By Max McClung

Background

- Most important attribute in an NBA player: Availability
- Being on the court comes before how you play on the court
- How can we better understand what makes a player miss games
- How does this affect NBA Executive/Coaching Decisions

An Overview

- NBA has struggled with Injury Prone Players
- Analyze the demographics that tend to be Injury Prone
- Utilize recent game data to create predictive model
- Visualize decision making process to see what contributes to availability

Dataset

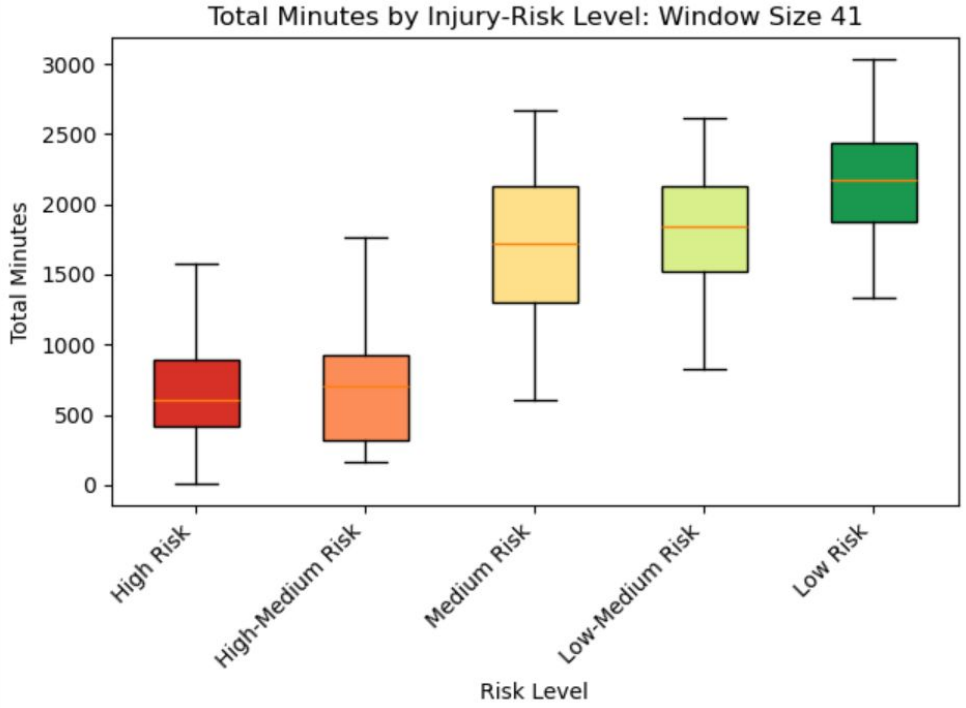
Player Name	Team	Seasons Played	Age	Position	Height (Inches)	Weight (lbs)	MPG	Games Played	Back to Backs	Total Minutes	Games Missed	82 Bit Wise Rep
Aaron Gordon	Nuggets	10	29	4	80	235	28.4	51	5	1448	31	1111111000000000
AJ Green	Bucks	2	25	2	76	190	22.7	73	15	1657	9	1111101111111111
Al Horford	Celtics	17	38	5	81	240	27.7	60	0	1662	22	1111101111110110
Alec Burks	Heat	13	33	2	77	214	17.6	49	7	862	33	0110111000001111
Alexandre Sarr	Wizards	0	20	4	84	205	27.1	67	10	1816	15	1111111111111111
Alperen Sengun	Rockets	3	22	5	82	243	31.5	76	15	2394	6	1111111111111111
Andre Drummond	Bulls	12	31	5	82	279	18.8	40	2	752	42	1111111111011111
Andrew Nembhard	Pacers	2	25	1	76	191	28.9	65	9	1879	17	1111101100000000

- 150 Current NBA Players
- 13 Features (2 Categorical, 10 Numerical, 1 Binary Representation)
- Workload (MPG, Games Played, etc.) and Player Demographics (Age, Weight, etc.)
- 82 bit number, representing every game of the season for each player

(References at end of Slideshow)

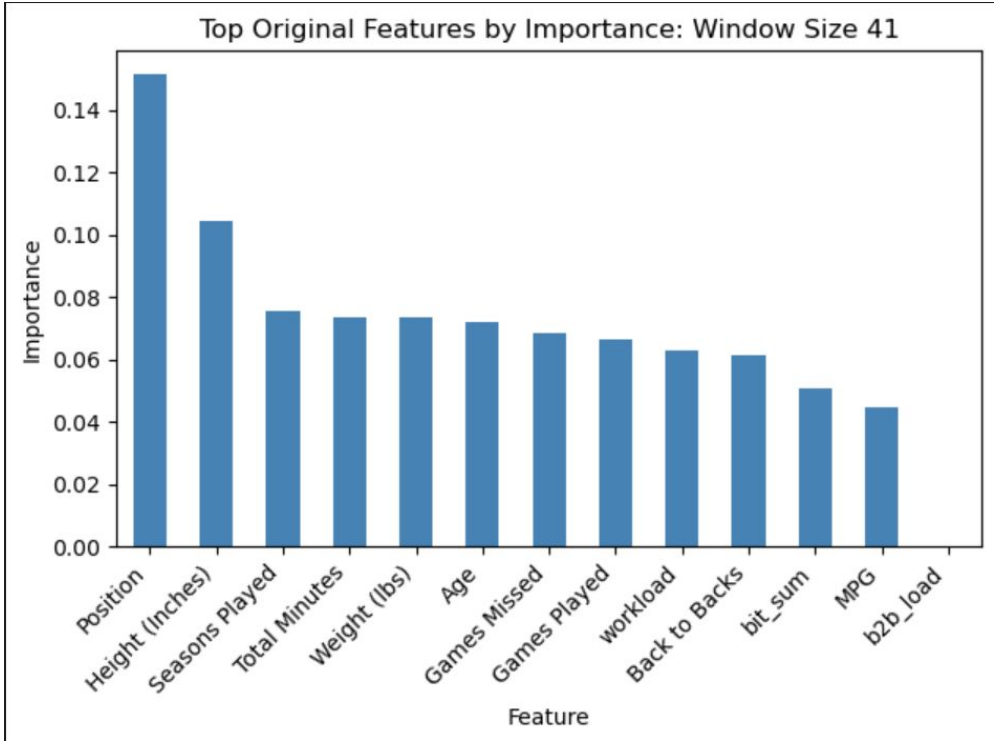
Data Improvement and What Contributes?

- Sliding Window Analysis
- K Means Grouping
- What Factors Contribute

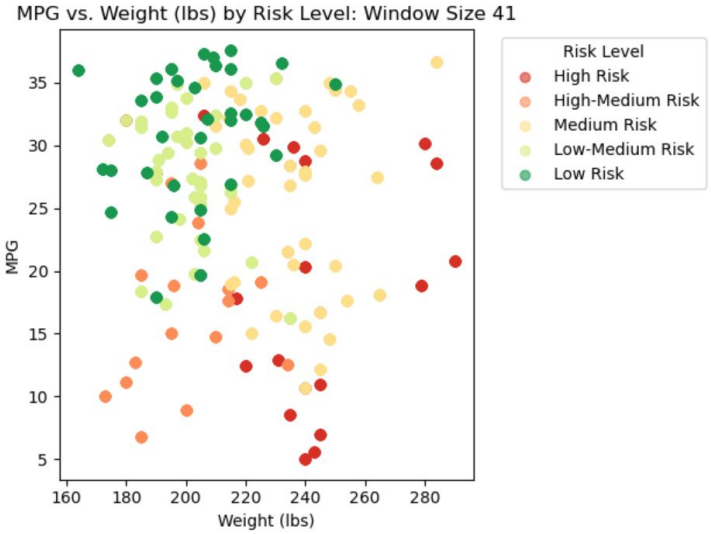


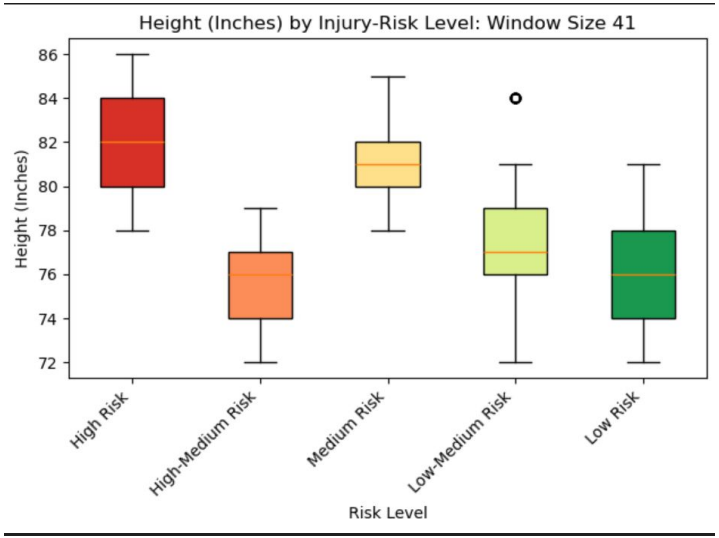
Additional Feature Engineering

- Calculated additional time-series features
- Workload, bit_sum, b2b_load
- Used to further improve model prediction and grouping



Interesting Finds

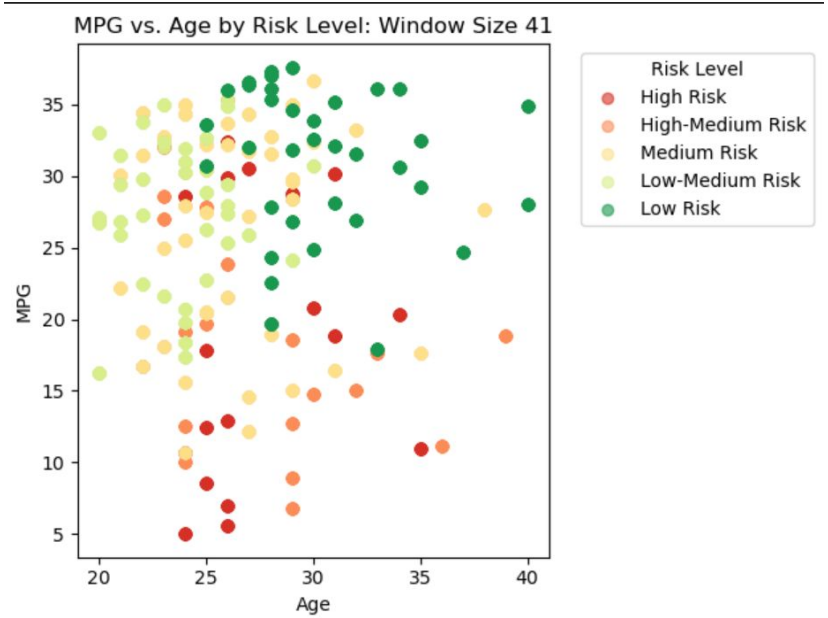




- Heavier + Taller Players Trend Risky
- Still see risk in lighter + shorter
- May affect Roster decisions

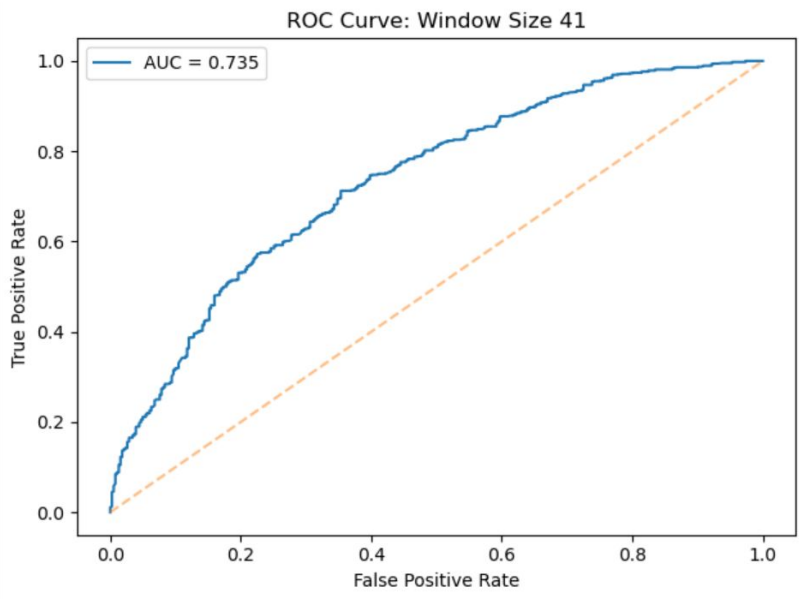
Interesting Finds

- 21-25 more risk than 25-30
- Likely due to college adjustments
- 40+ players seem low risk
- Limited data in this range

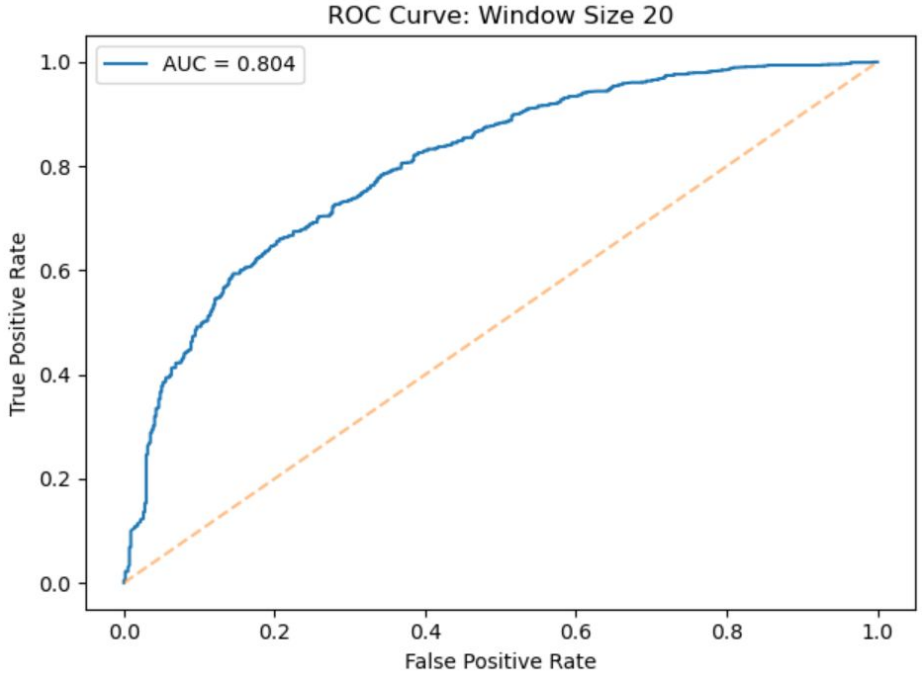


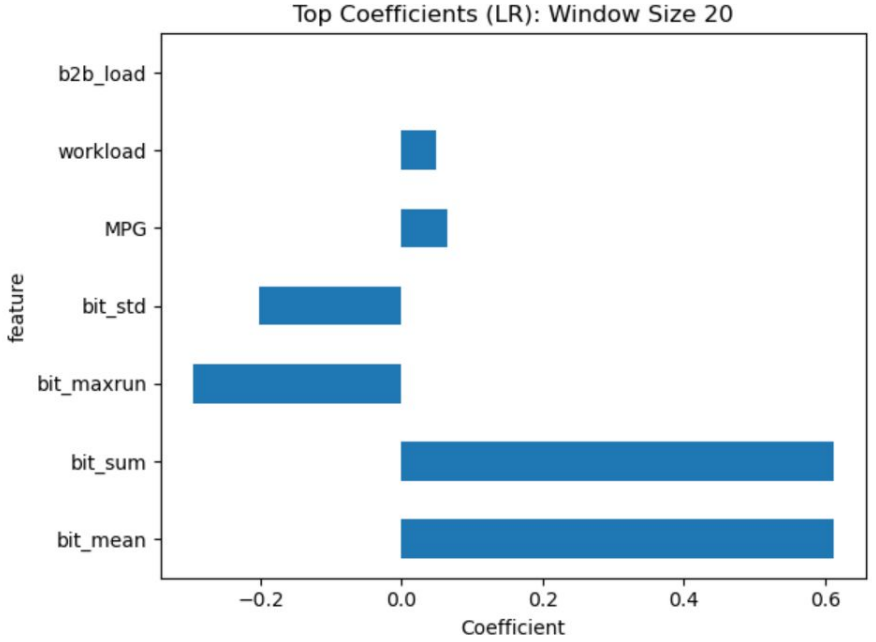
Predictive Model

- Predict whether a player will play or not
- Equal Weighting Logistic Regression
- Window Size too large
- Recent history better for prediction
- Shrink window size and retest



Logistic Regression: Window Size 20

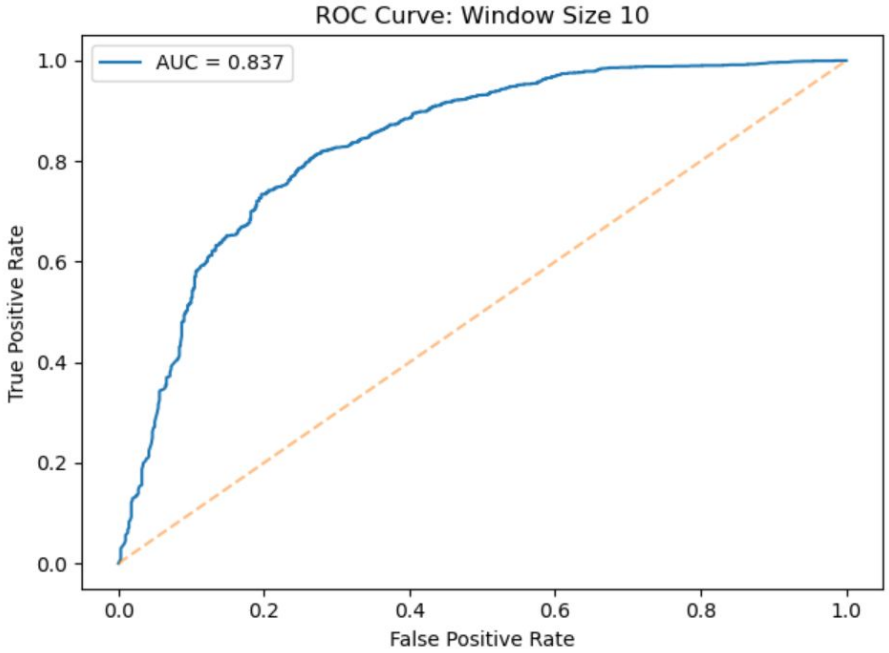




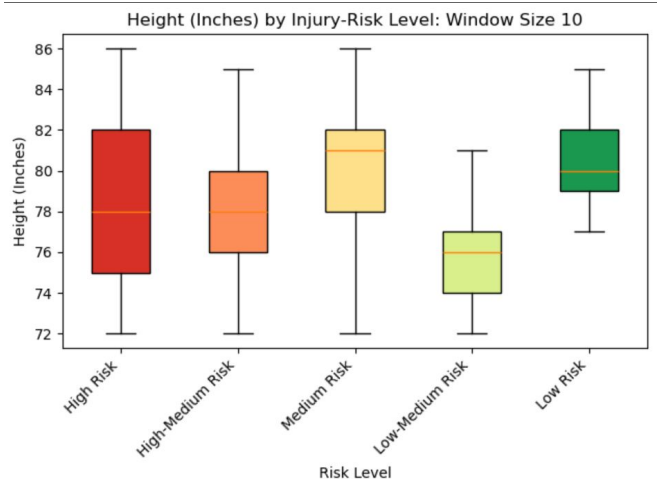
Logistic Regression: Window Size 10

```
Top positive drivers: Window Size %f {10}
    feature      coef
    bit_sum  0.762772
   bit_mean  0.762772
        MPG  0.085433
   b2b_load  0.000000
   workload -0.011477
    bit_std -0.196992
 bit_maxrun -0.332758
```

- Short term more accurate
- Recently Available trends Available
- Bit_maxrun shows opposite
- Long streaks, lead to risk
- Play often but not always



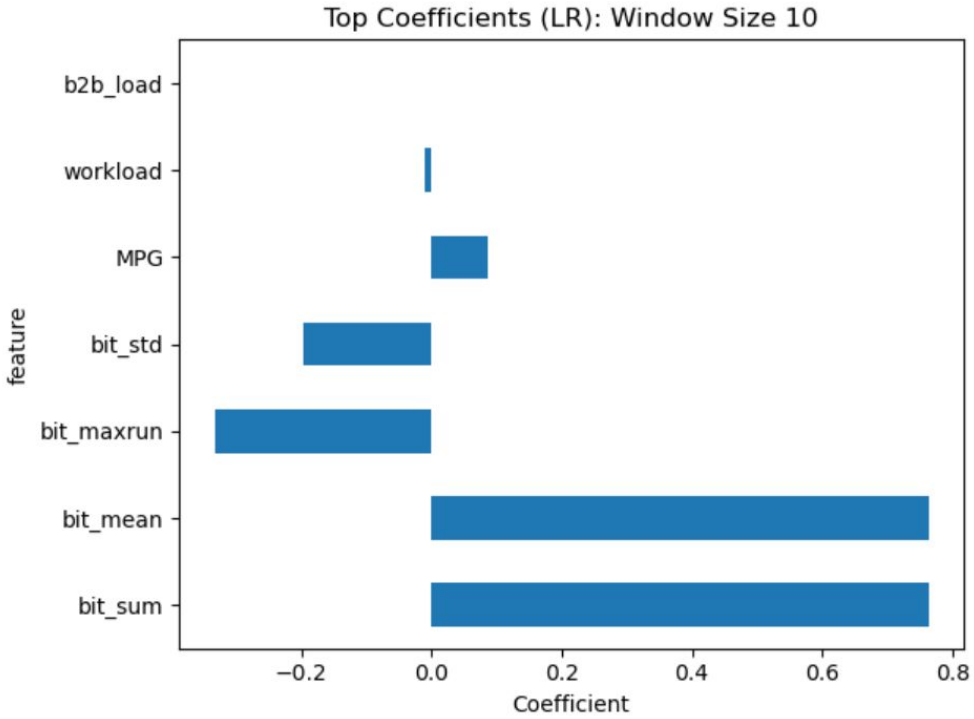
Window Size Importance



- Window Size affects amount of data
- Lower window size leads to large overfitting
- K-means and grouping shows lackluster conclusions
- Predictive model does grow more accurate
- Long window size to analyze injury demographics
- Short window size to predict players availability trends

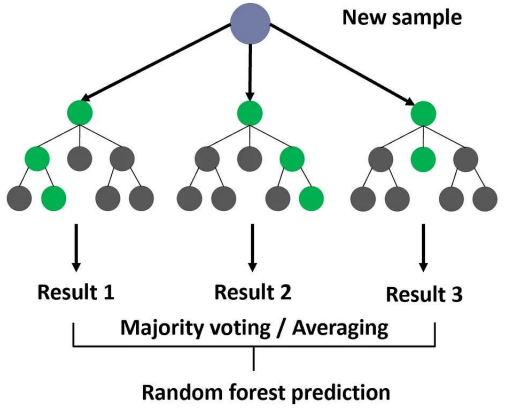
Logistic Regressions Limitations

- Linear effect of features
- Features interact
- Limited improvement
- Reach peak of prediction
- New model with RandomForest



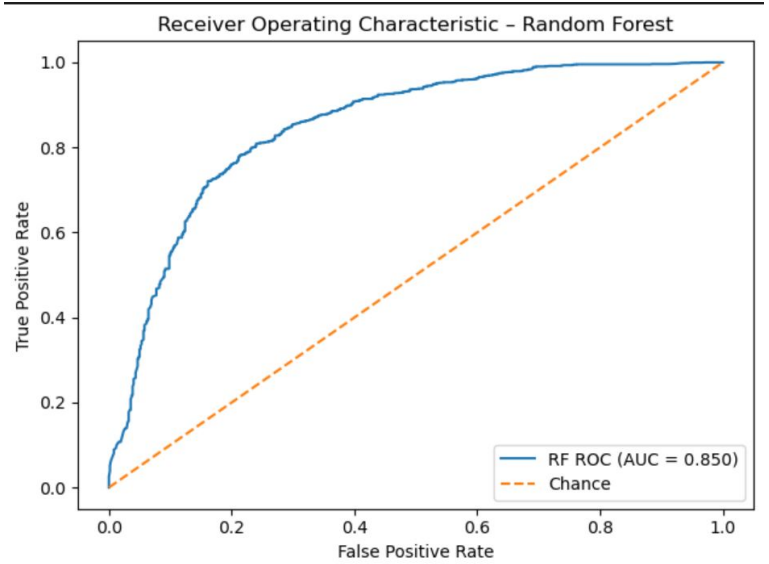
Random Forest Classifier

- Decision boundaries now non-linear
- Not as interpretable but better predictions
- Splits in paths lead to feature interactions
- Already know that recent data impacts more
- How does RandomForest at Window Size 10 perform



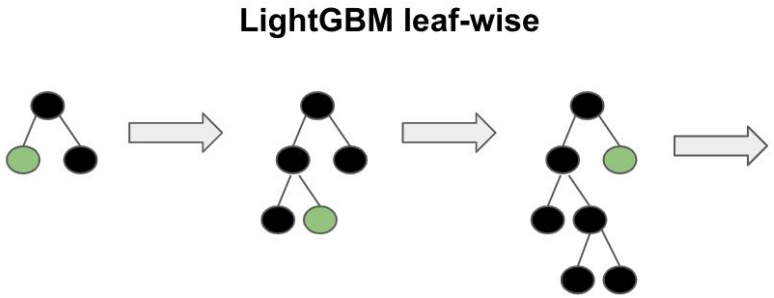
Random Forest Model

```
Classification report (RF):
              precision    recall  f1-score   support

           0      0.593     0.786     0.676       621
           1      0.901     0.782     0.837      1539

    accuracy                          0.783      2160
   macro avg      0.747     0.784     0.757      2160
weighted avg      0.812     0.783     0.791      2160

Test  AUC (RF):   0.8486715237428575
```

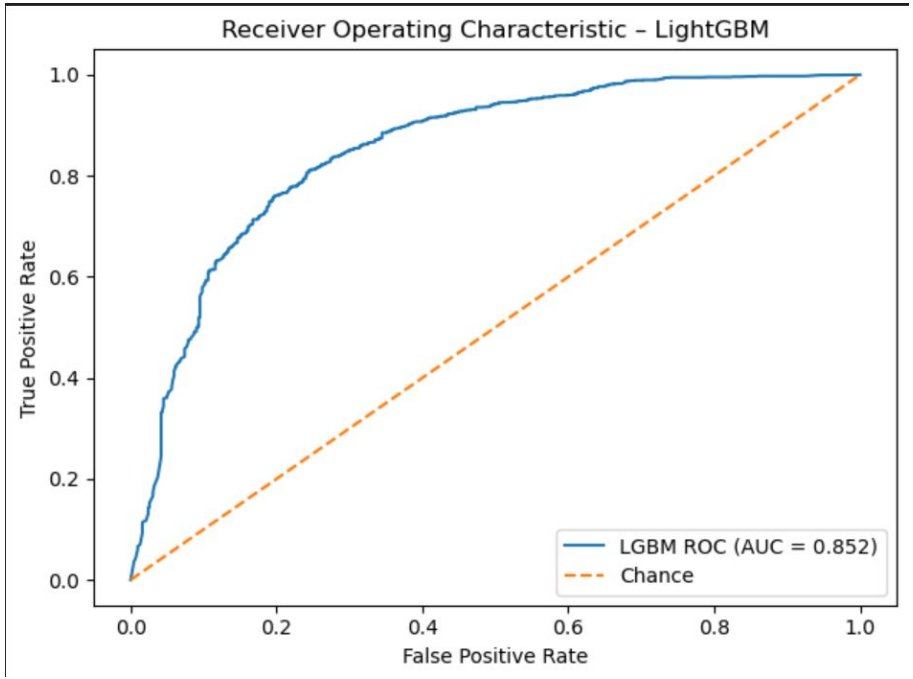


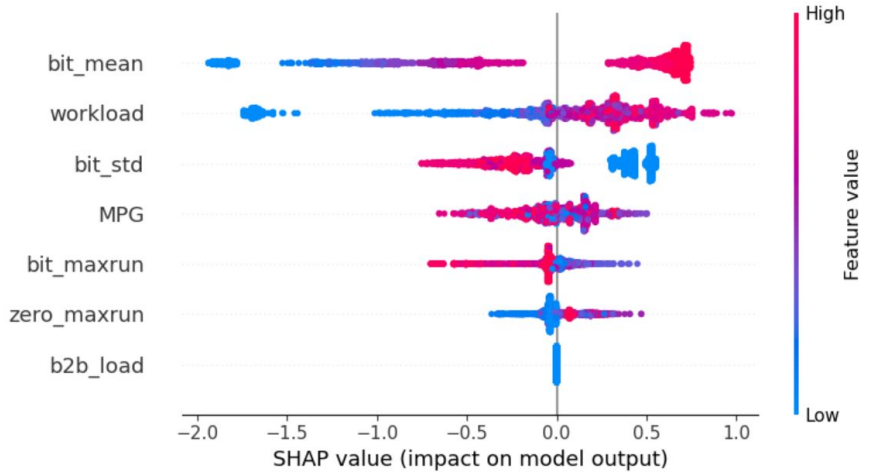
From Random Forest to LGBM Model

- Gradient boosting can remove errors
- Improved training time
- Can capture more complex feature interactions



LGBM Model


Receiver Operating Characteristic – LightGBM
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
LGBM ROC (AUC = 0.852)
Chance

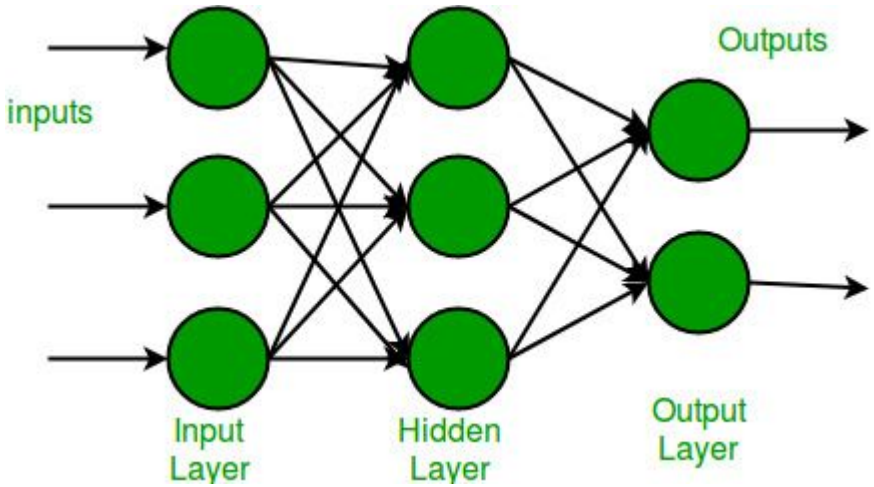

bit_mean
workload
bit_std
MPG
bit_maxrun
zero_maxrun
b2b_load
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
SHAP value (impact on model output)
High
Feature value
Low

MLP Model

- Approximate

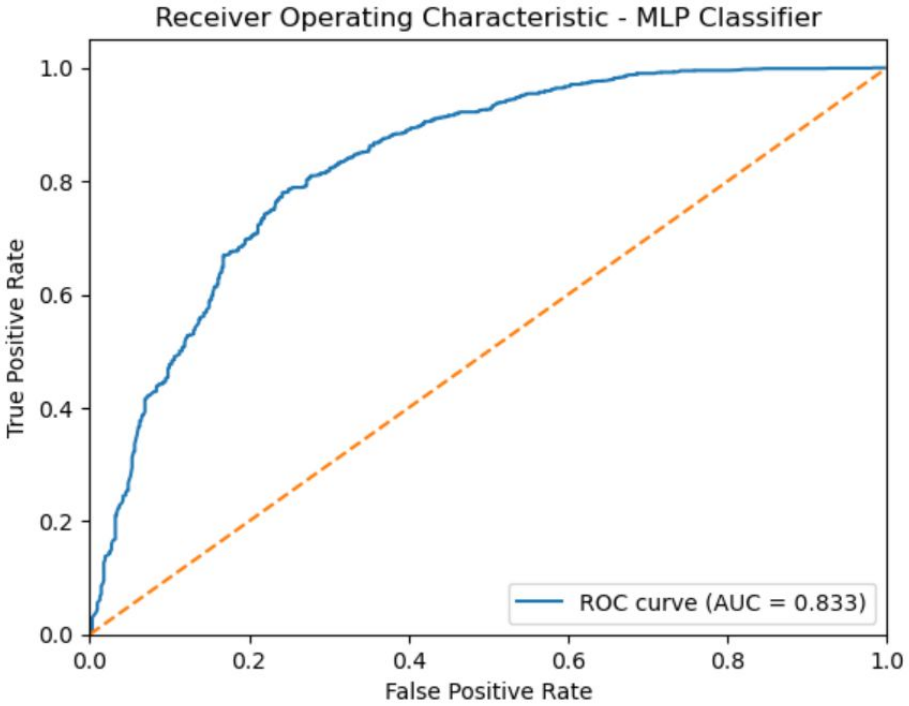
  Complex-Nonlinear Functions
- Input features affect throughout
- Same change in one input, not

  always same change in output



MLP Model AUC

- After tuning peaked lower than LGBM
- Dataset likely not large enough
- Trained across seasons would improve data prediction



Applications of Data

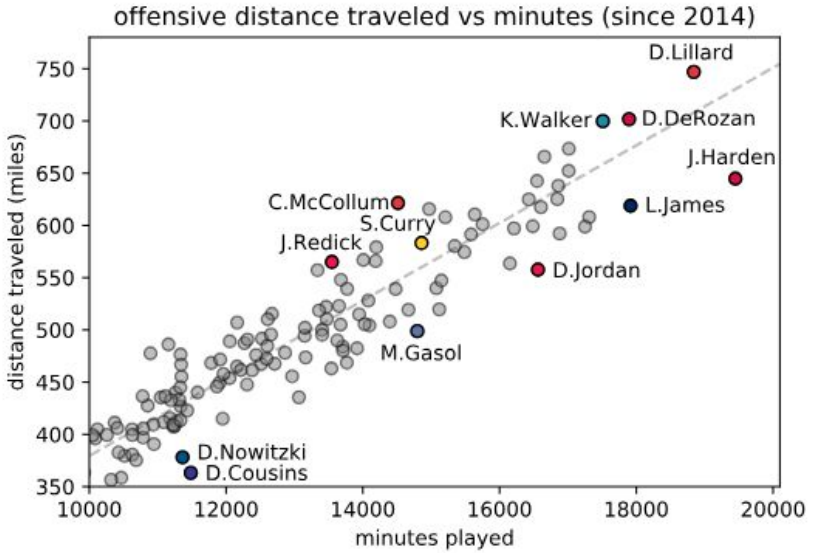
- Risk categories effects how executives might fill out a roster
- Taller and Heavier Players more risky
- 7 Big Men vs 5 Guards rather than vice versa
- If they get hurt more often, likely need more in case
- Predictive model can help keep players healthy
- If model is concerned about players likelihood to play, maybe rest

Data Limitations and Concerns

- Load management, players missing for non-”injuries”
- 1 day of “injury” not as impactful as more
- Lack of biometric data
- More in depth analysis could include injury type
- More seasons could help influence findings

Future of Research

- Much more to understand in world of injuries!
- Injuries are unpredictable, hard to know when someone will suffer one
- Utilize biometric and in depth data
- Attempt to prevent further injuries!
- Educate Executive Decisions



Why NBA Fans HATE this!

- Predicting an injury is almost impossible
- Resting a player with little to no reason becomes a slippery slope
- May keep a player healthy, but impossible to prove if you were “right”
- Fans travel to see games and can interfere with viewership
- Important to balance the analytics as well as the feel of the game
- Shouldn’t replace decision making, but help give coaches more detail!

Questions?
Thank you!

References

Basketball statistics & history of every team & NBA and WNBA players | Basketball-Reference.com. (n.d.). Basketball-Reference.com. https://www.basketball-reference.com/

StatMuse | Search StatMuse, save time. (n.d.). StatMuse. https://www.statmuse.com/

LandOfBasketball.com, information about NBA players, teams and championships. (n.d.). https://www.landofbasketball.com/index.htm

Official NBA Stats | Stats | NBA.Com. (n.d.). NBA. https://www.nba.com/stats

Lewis, M. (2018). It's a Hard-Knock life: game load, fatigue, and injury risk in the National Basketball Association. *Journal of Athletic Training*, *53*(5), 503–509. https://doi.org/10.4085/1062-6050-243-17