

Executive Summary – Running Hot: Determining the probability of injuries

Reed Ballesteros and Nadeem Patel

MSDS 356 Winter 2023

Northwestern University, Sports Performance Analytics

03/07/2022

To: Team owner, general manager, coaching staff, training staff, and medical staff
Re: "Running Hot" - Evaluating recent NBA Player performance in regards to injury risk

Hello,

Load management has become increasingly prevalent in the National Basketball Association (NBA), but it has faced criticism for potentially compromising the integrity of the regular season by negatively impacting team performance and fan engagement. However, former NBA player JJ Reddick argues that it is not always in the best interest of organizations to have their athletes playing every game. On his podcast, *The Old Man & the Three*, Reddick said, "A lot of times, there's a performance team, and they have all these different metrics measuring you on a weekly basis, and they call it 'running hot.' They will legitimately give you a heads up, and say, 'Hey, all your data says you're running hot. You may need to sit out next week. Which game looks better for you?'" (Reddick 2022).

While teams do not publicly report the metrics utilized to determine why their players might be asked to sit out, this study attempts to create a model using metrics that are focused on physical attributes and on-court activity since such factors tend to contribute toward injuries, according to previous studies. The "Running Hot Engine" can take in a player's recent game log, along with their game log data from the week before, as well as career injury information. Based on a given threshold, the model can evaluate if a player is "running hot" and recommends whether or not a player should sit out the next game on the schedule. With limited time for prototyping, game log data of over 100 players from the 2021-22 NBA regular season who averaged more than 27 minutes per game were sourced. This resulted in a dataset of over 6,699 individual game logs used to train and test our model.

Based on model analysis, the resulting probability predictor prototype used in the current iteration of the Running Hot Engine is a logistic regression model that ingests 42 features of

player's game log to calculate their "Running Hot Index," which is a value from 0 (low) to 100 (high). The "Running Hot Threshold" has a default value of 50, but it can be set to any value from 0 to 100 as needed. If the index exceeds the threshold, the engine will flag a player as "running hot" as a recommendation for them to sit out the next game. For example, the game log of Dallas Maverick Luka Dončić's second-to-last game of the 2021-22 season can be fed into the Running Hot Engine:

SEASON_YEAR	2021-22	POSS	62
PLAYER_ID	1629029	PTS_PAINT	10
PLAYER_NAME	Luka Dončić	PCT_PTS_2PT_MR	0.0
TEAM_ID	1610612742	AGE	23
TEAM_ABBREVIATION	DAL	PLAYER_HEIGHT_INCHES	79
TEAM_NAME	Dallas Mavericks	PLAYER_WEIGHT	230
GAME_ID	22101209	FGM_PAINT	5
GAME_DATE	2022-04-08T00:00:00	FGM_2PT_MR	0
MATCHUP	DAL vs. POR	PW_MIN	109.3
WL	W	PW_GAMES	3
MIN	29.6	PW_FG3A	29
FGM	12	PW_FG2A	36
FGA	21	PW_OREB	1
FG3M	7	PW_DREB	22
FG3A	14	PW_ASTS	35
FTM	8	PW_TOV	17
FTA	11	PW_STL	4
OREB	0	PW_BLK	1
DREB	11	PW_POS	222
AST	7	PW_PF	9
TOV	5	PW_PFD	23
STL	0	PW_FGM_PAINT	14.0
BLK	1	PW_FGM_2PT_MR	5.0
PF	0	PW_PIE_AVG	0.160333
PFD	7	PW_USG_PCT_AVG	0.417
PTS	39	CAREER_INJURIES	26
USG_PCT	0.443	DAYS_LAST_INJURY	18
PIE	0.34	OUT_NEXT_GAME	0

For Doncic's sample data, the Running Hot Engine gives the following evaluation:

PLAYER_NAME	Luka Doncic
GAME_DATE	2022-04-08T00:00:00
RUNNING_HOT_IDX	8.7
RUNNING_HOT	False
OUT_NEXT_GAME	0

Based on the results, Doncic's Running Hot Index was 8.7, which is below the given Running Hot Threshold of 50. This means Luka was not "running hot" from his April 8, 2022 game against the Portland Trailblazers and recommends him to play the next game on the Maverick's schedule, the last regular season game against the San Antonio Spurs on April 10, 2022. He did play in that game, logging 28 minutes and scoring 26 points.

Analysis can be performed using previous game data to check if the Running Hot Engine would have determined if a player was "running hot" but still played the next game on the team schedule. Using all of Doncic's 2021-22 regular season game logs and the default threshold of 50, the analysis gives the following:

```
TOTAL GAME LOGS: 65
TOTAL RUNNING HOT IN GAME LOGS: 10
RUNNING_HOT/OUT_NEXT_GAME MATCH PERCENTAGE: 90.77%
TIMES SAT OUT NEXT GAME WHILE ABOVE RUNNING HOT THRESHOLD (50) (RECOMMENDED BY MODEL): 5
TIMES PLAYED NEXT GAME WHILE UNDER RUNNING HOT THRESHOLD (50) (RECOMMENDED BY MODEL): 54
TIMES PLAYED NEXT GAME WHILE ABOVE RUNNING HOT THRESHOLD (50) (*NOT* RECOMMENDED BY MODEL): 5
TIMES SAT NEXT GAME WHILE UNDER RUNNING HOT THRESHOLD (50) (*NOT* RECOMMENDED BY MODEL): 1
```

Using the current build of the Running Hot Engine, the analysis showed that Doncic was "running hot" going into five games he played in the regular season. However, he did not suffer an injury in those games and even had very solid performances. A narrative behind this analysis could show that despite the risk for injury, a player like Doncic continued through the NBA season, which can be grueling, and in doing so, was able to deliver solid performances in those potentially risky games.

Given the findings, the Running Hot Engine has limits at its current state. There is technical debt in regard to data collection, and if resolved would result in a much larger and richer dataset of game logs from more players and seasons that can be used in fitting the model. A larger dataset can possibly help mitigate bias and overfitting issues found in the current state of the engine. Future iterations would attempt to integrate tracking and biometric data such as speed, distance, and heart rate, and other features that could potentially enhance the dataset further (we will be very mindful of HIPAA rules in terms of what we can or cannot use in the dataset).

The current iteration of the Running Hot Engine should be considered more as a proof of concept, and with time, can have the potential to be a complementary tool to help the team's coaching, training, and medical staff in their decision-making when it comes to resting their players. Thank you for your time.

Best,

Reed Ballesteros and Nadeem Patel